GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 5, 2006, 19:12:00 ; Search time 2902 Seconds
(without alignments)
10479.410 Million cell updates/sec

Title: US-10-664-705-145
Perfect score: 535
Sequence: 1 cggccgcctcctggcccag........cgcctgcgctcggtgctcat 535

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	535	100.0	928	9	RATCBR3	M19262 Rat clathri
2	471	88.0	982	6	AX401731	AX401731 Sequence
3	471	88.0	982	6	AX827493	AX827493 Sequence
4	471	88.0	982	9	RATCBR2	M15883 Rat clathri
5	451.8	84.4	636	8	BT007028	BT007028 Homo sapi
6	451.8	84.4	636	8	CRS36577	CR536577 Homo sapi
7	451.8	84.4	636	11	AY889907	AY889907 Synthetic
8	451.8	84.4	636	11	AY892384	AY892384 Synthetic
9	451.8	84.4	636	11	AY892385	AY892385 Synthetic
10	451.8	84.4	636	11	BT007759	BT007759 Synthetic
11	451.8	84.4	1051	8	HUMLCTHB	M20470 Human lymph
12	451.8	84.4	1099	8	BC006457	BC006457 Homo sapi
13	451.8	84.4	1164	8	BC006332	BC006332 Homo sapi
14	442.2	82.7	1009	9	BC070404	BC070404 Mus muscu
15	429.4	80.3	1033	4	BTLCBRL	X04853 Bos taurus
16	387.8	72.5	1134	8	HUMBCTHB	M20469 Human brain
17	365.4	68.3	1029	4	BTLCBRB	X04852 Bos taurus
18	325	60.7	688	6	CQ729113	CQ729113 Sequence
19	323.8	60.5	2615	5	AJ720113	AJ720113 Gallus ga
20	261	48.8	738	6	CQ729572	CQ729572 Sequence
21	239.8	44.8	2317	5	BC084464	BC084464 Xenopus t
22	239.8	44.8	2773	5	CR761042	CR761042 Xenopus t
23	236.8	44.3	586	9	BC037505	BC037505 Mus muscu
24	235	43.9	1711	5	BC060412	BC060412 Xenopus l
25	224	41.9	1751	5	BC077312	BC077312 Xenopus l
26	223	41.7	1005	4	X04851	X04851 Bos taurus
27	219.8	41.1	1058	9	BC057660	BC057660 Mus muscu
28	216.6	40.5	997	5	AJ721067	AJ721067 Gallus ga
29	215	40.2	1034	9	RATCAR3	M19261 Rat clathri
30	215	40.2	1106	9	BC087577	BC087577 Rattus no
31	213.4	39.9	657	8	BT007170	BT007170 Homo sapi
32	213.4	39.9	657	11	AY890294	AY890294 Synthetic
33	213.4	39.9	657	11	AY890295	AY890295 Synthetic
34	213.4	39.9	657	11	AY892771	AY892771 Synthetic
35	213.4	39.9	657	11	BT008288	BT008288 Synthetic
36	213.4	39.9	1023	6	CQ724687	CQ724687 Sequence
37	213.4	39.9	1023	8	HUMLCTHA	M20472 Human lymph
38	213.4	39.9	1055	8	BC019287	BC019287 Homo sapi
39	213.4	39.9	1103	8	BC009201	BC009201 Homo sapi
40	197.4	36.9	1012	5	BC092127	BC092127 Xenopus l
41	187.8	35.1	927	5	CR760072	CR760072 Xenopus t
42	187.8	35.1	1072	5	BC061272	BC061272 Xenopus t
43	176.2	32.9	958	6	AX305857	AX305857 Sequence
44	174.2	32.6	400	6	CQ665260	CQ665260 Sequence
45	152.2	28.4	1329	5	BC060939	BC060939 Danio rer

ALIGNMENTS

RESULT 1
RATCBR3
LOCUS   RATCBR3   928 bp   mRNA   linear   ROD 27-APR-1993
DEFINITION Rat clathrin light chain (LCB3) mRNA, complete cds.
ACCESSION   M19262
VERSION   M19262.1 GI:203360
KEYWORDS   alternative splicing; clathrin.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 928)
AUTHORS   Kirchhausen,T., Scarmato,P., Harrison,S.C., Monroe,J.J., Chow,E.P.,
            Mattaliano,R.J., Ramachandran,K.L., Smart,J.B., Ahn,A.H. and
            Brosius,J.
TITLE   Clathrin light chains LCA and LCB are similar, polymorphic, and
            share repeated heptad motifs
JOURNAL   Science 236 (4799), 320-324 (1987)
PUBMED   3563513
COMMENT   Original source text: Rat liver, cDNA to mRNA, clones
            lambda[4-1,7-1] and brain, cDNA to mRNA, clone lambda-TWC1-3.
            Draft entry and computer-readable sequence for [1] kindly provided
            by T.Kirchhausen, 05-OCT-1987.
FEATURES   Location/Qualifiers
     source   1..928
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
     mRNA   <1..928
                     /product="LCB3 mRNA"
     CDS   29..664
                     /note="clathryn light chain (LCB3)"
                     /codon_start=1
                     /protein_id="AAA40891.1"
                     /db_xref="GI:203361"
                     /translation="MAEDFGFFSSSESGAPEAAEBDPAAAFLAQQESEIAGIENDSGF
                     GAPASQVASAQPGLASGGGSEDMGTTVNGDVFQEANGPADGYAAIAQADRLTQSPES
                     IRKWREEQKRLQELDAASKVTEQEWREKAKKDLEEWNQRQSEQVEKNKINNRASBEA
                     FVKESKEETPGTEWEKVAQLCDFNPKSSKQCKDVSRLRSVLMSLKQTPLSR"
ORIGIN   74 bp upstream of AvaI site.

Left column (continuation of RESULT 1)

```
Query Match          100.0%;  Score 535;  DB 9;  Length 928;
Best Local Similarity 100.0%;  Pred. No. 2.4e-111;
Matches  535;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 CGGCCGCCTTCCTGGCCGCGCAGGAGAGGAGCGAGGATTGCTGGCATCGAGAATGACTCGGGTT  60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   99 CGGCCGCCTTCCTGGCCGCGCAGGAGAGGAGCGAGGATTGCTGGCATCGAGAATGACTCGGGTT 158

Qy   61 TCGGGGCACCTGCCCGCGCCCAGGTGGCCTCTGCCGCCCGGACTCGCGAGCGGGGGTG 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  159 TCGGGGCACCTGCCCGCGCCCAGGTGGCCTCTGCCGCCCGGACTCGCGAGCGGGGGTG 218

Qy  121 GTTCGGAGGACATGGGGACTACAGTCAATGGCGGACAGGTTGACTCAGGAGCTCAGGAGCATCC 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  219 GTTCGGAGGACATGGGGACTACAGTCAATGGCGGACAGGTTGACTCAGGAGCTCAGGAGCATCC 278

Qy  181 CCGATGGCTACGTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCTCAGGAGCCCTG 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  279 CCGATGGCTACGTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCTCAGGAGCCCTG 338

Qy  241 GCAAGTGGAGAGGAGCAGAAGGCTGGATGGTTGGATGCTGCTCGTCGAAGGTGA 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  339 GCAAGTGGAGAGGAGCAGAAGGCTGGATGGTTGGATGCTGCTCGTCGAAGGTGA 398

Qy  301 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAGACCTGGAACCAGCGCCAAA 360
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  399 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAGACCTGGAACCAGCGCCAAA 458

Qy  361 GCAAGTGGAGAGGAGCAGAAGGCTGCCAGGAGTTGCCTCGAAGGTGA 398
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  459 GCAAGTGGAGAGGAGCAGAAGGCTGCCAGGAGTTGCCTCGAAGGTGA 518
```

Right column (continuation of RESULT 2)

```
Db   99 CGGCCGCCTTCCTGGCCGCGCAGGAGAGGAGCGAGGATTGCTGGCATCGAGAATGACTCGGGTT 158
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   61 TCGGGGCACCTGCCCGCGCCCAGGTGGCCTCTGCCGCCCGGACTCGCGAGCGGGGGTG 120

Db  159 TCGGGGCACCTGCCCGCGCCCAGGTGGCCTCTGCCGCCCGGACTCGCGAGCGGGGGTG 218
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  121 GTTCGGAGGACATGGGGACTACAGTCAATGGCGGACAGGTTGACTCAGGAGCTCAGGAGCATCC 180

Db  219 GTTCGGAGGACATGGGGACTACAGTCAATGGCGGACAGGTTGACTCAGGAGCTCAGGAGCATCC 278
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  181 CCGATGGCTACGTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCTCAGGAGCCCTG 240

Db  279 CCGATGGCTACGTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCTCAGGAGCCCTG 338
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  241 GCAAGTGGAGAGGAGCAGAAGGCTGGATGGTTGGATGCTGCTCGTCGAAGGTGA 300

Db  339 GCAAGTGGAGAGGAGCAGAAGGCTGGATGGTTGGATGCTGCTCGTCGAAGGTGA 398
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  301 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAGACCTGGAACCAGCGCCAAA 360

Db  399 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAGACCTGGAACCAGCGCCAAA 458
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  361 GTGAACAGGTTGAGAAGAACAAGATCAACAACAG 394

Db  459 GTGAACAGGTTGAGAAGAACAAGATCAACAACAGATCGCTGACAAAGCGTTCTACCAGC 518

Db  395 ------GGCATCGGAAGAGGCGTTTTGTGAAAGAATCCA 426
              ||||||||||||||||||||||||||||||||
Qy  519 AGCCAGATGCTGATACCATTGGCTATGTTGGCATCGGAAGAGGCGTTTTGTGAAAGAATCCA 578

Db  519 AGGAGGAGACCCCCAGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTA 486
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  427 AGGAGGAGACCCCCAGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTA 486

Db  579 AGGAGGCAGCAAGCAATGTGTCCCGCCTGCGTCGGTGCTCAT 638
        ||||||||||||||||||||||||||||||||||||||||||
Qy  487 AGAGCAGCAAGCAATGTGTCCCGCCTGCGTCGGTGCTCAT 535

Db  639 AGAGCAGCAAGCAATGTGTAAAGACGTGTCCCGCCTGCGTCGGTGCTCAT 687
```

RESULT 2

```
AX401731

LOCUS       AX401731                 982 bp    DNA     linear   PAT 06-JUN-2002
DEFINITION  Sequence 1407 from Patent WO0210453.
ACCESSION   AX401731
VERSION     AX401731.1  GI:21337911
KEYWORDS    .
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1
  AUTHORS   Mendrick,D., Porter,M.W., Johnson,K.R., Castle,A.L. and
            Elashoff,M.R.
  TITLE     Molecular toxicology modeling
  JOURNAL   Patent: WO 0210453-A 1407 07-FEB-2002;
            Gene Logic, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..982
                     /organism="Rattus norvegicus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10116"
                     /note="EMBL/GenBank Accession No. M15883"
ORIGIN

Query Match           88.0%;  Score 471;  DB 6;  Length 982;
Best Local Similarity 90.8%;  Pred. No. 8.8e-97;
Matches  535;  Conservative   0;  Mismatches   0;  Indels  54;  Gaps   1;

Qy    1 CGGCCGCCTTCCTGGCCGCGCAGGAGAGGAGCGAGGATTGCTGGCATCGAGAATGACTCGGGTT 60
```

RESULT 3

```
AX827493

LOCUS       AX827493                 982 bp    DNA     linear   PAT 12-DEC-2003
DEFINITION  Sequence 227 from Patent EP1344834.
ACCESSION   AX827493
VERSION     AX827493.1  GI:39837681
KEYWORDS    .
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1
  AUTHORS   Boess,F., Suter-Dick,L. and Wolf,D.
  TITLE     Methods for the toxicity prediction of a compound
  JOURNAL   Patent: EP 1344834-A 227 17-SEP-2003;
            F. HOFFMANN-LA ROCHE AG (CH)
FEATURES             Location/Qualifiers
     source          1..982
                     /organism="Rattus norvegicus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10116"
ORIGIN

Query Match           88.0%;  Score 471;  DB 6;  Length 982;
Best Local Similarity 90.8%;  Pred. No. 8.8e-97;
Matches  535;  Conservative   0;  Mismatches   0;  Indels  54;  Gaps   1;

Qy    1 CGGCCGCCTTCCTGGCCGCGCAGGAGAGGAGCGAGGATTGCTGGCATCGAGAATGACTCGGGTT 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   99 CGGCCGCCTTCCTGGCCGCGCAGGAGAGGAGCGAGGATTGCTGGCATCGAGAATGACTCGGGTT 158

Qy   61 TCGGGGCACCTGCCCGCGCCCAGGTGGCCTCTGCCGCCCGGACTCGCGAGCGGGGGTG 120
```

GAPAASQVASAQPGLASGGGSBDMGTTVNGDVFQBANGPADGYAAIAQADRLTQEPES
IRKWREBQKKRIQELDAASKVTEQEWREKAKKDLEEWNQROSEQVEKNKINNRIADKA
FYQQPDADTIGVASEEAFVKBSKEETPGTEWEKVAQLCDFNPKSSKQCKDVSRLRSV
LMSLKQTPLSR"

ORIGIN   74 bp upstream of AvaI site.

Query Match   88.0%; Score 471; DB 9; Length 982;
Best Local Similarity 90.8%; Pred. No. 8.8e-97;
Matches 535; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

Qy   1 CGGCCGCCTTCCTGCCCAGGAGGAGCCGAGAATTGCTGCATCGAGAATGACTCGGGTT 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   99 CGGGCCGCCCTTCCTGCCCCACCAGGAGGAGCCGAGAATTGCTGCATCGAGAATGACTCGGGTT 158

Qy   61 TCGGGGCCACCTGGCCTCTGCGCCCAGGTGGCCTCTGCGCCCGCGGACTCGCGAGCGGGGGTG 120
            |||||||||||||||||||||||||||||||||||||||||||||||
Db   159 TCGGGGCCACCTGGCCTCTGCGCCCAGGTGGCCTCTGCGCCCGCGGACTCGCGAGCGGGGGTG 218

Qy   121 GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTCAGGAGGCTAACGGGCCTG 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   219 GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTCAGGAGGCTAACGGGCCTG 278

Qy   181 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCCTGAGAGCATCC 240
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db   279 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCCTGAGAGCATCC 338

Qy   241 GCAAGTGGAGAGAGGAGCAGAAGAAAGGCTGCAGGAGTGTGCAGGAGGTTGGATGCGTCGGAGGTGA 300
            |||||||||||||||||||||||||||||||||||||||||||||||
Db   339 GCAAGTGGAGAGAGGAGCAGAAGAAAGGCTGCAGGAGTGTGCAGGAGGTTGGATGCGTCGGAGGTGA 398

Qy   301 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 360
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db   399 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 458

Qy   361 GTGAACAGGTTGAGAAGAACAAGATCAACAACAG------------------------ 394
            |||||||||||||||||||||||
Db   459 GTGAACAGGTTGAGAAGAACAAGATCAACAACACAGGATCGCTGACAAAGGCGTTCTACCAGC 518

Qy   395 --------GGCATCGGAAGAGGCTTTTGTGAAAGAATCCA 426
            ||||||||||||||||||||||||
Db   519 AGCCAGATGCTGATACCATTGGCTATGTGGCATCGGAAGAGGCTTTTGTGAAAGAATCCA 578

Qy   427 AGGAGGAGCAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTA 486
            |||||||||||||||||||||||||||||||||||||||
Db   579 AGGAGGAGCAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTA 638

Qy   487 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTCGGTGCTCAT 535
            |||||||||||||||||||||||||||||
Db   639 AGAGCAGCAAGCAATGTAAAGACGTGTCCCCGCCTCGGTCGCGTGCTCAT 687

RESULT 5
BT007028
LOCUS   BT007028   636 bp   mRNA   linear   PRI 13-MAY-2003
DEFINITION Homo sapiens clathrin, light polypeptide (Lcb) mRNA, complete cds.
ACCESSION   BT007028
VERSION   BT007028.1 GI:30582894
KEYWORDS   FLI_CDNA.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 636)
AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
TITLE   Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 636)
AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,

Db   159 TCGGGGCCACCTGGCCTCTGCGCCCAGGTGTTCGCGAGATGTGTTCGCGAGCCGGAGCTAACGGGCCTG 218
Qy   121 GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTCAGGAGGCTAACGGGCCTG 180
Db   219 GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTCAGGAGGCTAACGGGCCTG 278
Qy   181 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCCTGAGAGCATCC 240
Db   279 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCCTGAGAGCATCC 338
Qy   241 GCAAGTGGAGAGAGGAGCAGAAGAAAGGCTGCAGGAGTGTGCAGGAGGTTGGATGCGTCGGAGGTGA 300
Db   339 GCAAGTGGAGAGAGGAGCAGAAGAAAGGCTGCAGGAGTGTGCAGGAGGTTGGATGCGTCGGAGGTGA 398
Qy   301 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 360
Db   399 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 458
Qy   361 GTGAACAGGTTGAGAAGAACAAGATCAACAACAG------------------ 394
Db   459 GTGAACAGGTTGAGAAGAACAAGATCAACAACACAGGATCGCTGACAAAGGCGTTCTACCAGC 518
Qy   395 --------GGCATCGGAAGAGGCTTTTGTGAAAGAATCCA 426
Db   519 AGCCAGATGCTGATACCATTGGCTATGTGGCATCGGAAGAGGCTTTTGTGAAAGAATCCA 578
Qy   427 AGGAGGAGCAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTA 486
Db   579 AGGAGGAGCAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTA 638
Qy   487 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTCGGTGCTCAT 535
Db   639 AGAGCAGCAAGCAATGTAAAGACGTGTCCCCGCCTCGGTCGCGTGCTCAT 687

RESULT 4
RATCBR2
LOCUS   RATCBR2   982 bp   mRNA   linear   ROD 27-APR-1993
DEFINITION Rat clathrin light chain (LCB2) mRNA, complete cds.
ACCESSION   M15883
VERSION   M15883.1 GI:203358
KEYWORDS   alternative splicing; clathrin.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 982)
AUTHORS   Kirchhausen,T., Scarmato,P., Harrison,S.C., Monroe,J.J., Chow,B.P.,
            Mattaliano,R.J., Ramachandran,K.L., Smart,J.E., Ahn,A.H. and
            Brosius,J.
TITLE   Clathrin light chains LCA and LCB are similar, polymorphic, and
            share repeated heptad motifs
JOURNAL   Science 236 (4799), 320-324 (1987)
PUBMED   3563513
COMMENT   Original source text: Rat brain, cDNA to mRNA, clones pPPb5-2,
            pVI2Ab, lambda-ACS2-1, and lambda-BC8-4.
            Draft entry and computer-readable sequence for [1] kindly provided
            by T.Kirchhausen, 05-OCT-1987.
FEATURES   Location/Qualifiers
     source   1..982
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
     mRNA   <1..982
                     /product="LCB2 mRNA"
     CDS   29..718
                     /codon_start=1
                     /note="clathryn light chain (LCB2)"
                     /protein_id="AAA40890.1"
                     /db_xref="GI:203359"
                     /translation="MAEDFGFFSSESGAPEAAREEDPAAAFLAQQESEIAGIENDSGF

```
TITLE     Phelan,M. and Farmer,A.
JOURNAL   Direct Submission
          Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
          Circle, Palo Alto, CA 94303, USA
COMMENT   This CDS clone is a part of a collection of human full length
          expression clones generated by BD Biosciences Clontech and the
          Harvard Institute of Proteomics. Each CDS has been cloned in two
          forms: with and without stop-codon (to allow fusion with C-terminal
          tag). The CDS has been directionally cloned using BD In-Fusion(TM)
          cloning system between the SalI and HindIII sites of the pDNR-DUAL
          vector. Additional sequences in the clone: 'ACC' after SalI site
          and before 'ATG' to provide Kozak consensus sequence; 'GG' after
          last codon and before HindIII site to maintain reading frame.
          Clone distribution: http://bioinfo.clontech.com/orfclones
FEATURES             Location/Qualifiers
     source          1..636
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GH00332X1.0"
                     /clone_lib="BD Creator(TM) CDS Library derived from MGC
                     collection"
                     /lab_host="DH5alpha T1 resistant"
                     /note="Vector: pDNR-Dual"
     CDS             1..636
                     /codon_start=1
                     /product="clathrin, light polypeptide (Lcb)"
                     /protein_id="AAP35675.1"
                     /db_xref="GI:30582895"
                     /translation="MADDFGFFSSSESGAPEAAEEDPAAAFLAQQESEIAGIENDEGF
                     GAPAGSHAAPAQPGTSGAGSEDMGTTVNGDVFQEANGPADGYAAIAQADRLTOEPES
                     IRKWREEQRKRLQELDAASKVTEQEWREKAKKDLEEWNQRQSEQEVEKNKINNRASEEA
                     FVKSSKEETPGTEWEKVAQLCDFNPKSSKQCKDVSRLRSVLMSLKQTPLSR"
ORIGIN
Query Match  84.4%;  Score 451.8;  DB 8;  Length 636;
Best Local Similarity  90.3%;  Pred. No. 2e-92;
Matches  483;  Conservative  0;  Mismatches  52;  Indels  0;  Gaps  0;

Qy      1 CGGCCGCCTTCCTGGCCCAGCAGGAGAGCGAGATTCGTGGCATCGAGAATGACTCGGGTT   60
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     71 CGGCCCGCCTTCCTGGCCCAGCAGGAGAGCGAGATTCGTGGCATCGAGAACGACGAGGGCT  130

Qy     61 TCGGGGACCACCTGCCGCAGCGTGGCCTCTGCGCAGCCCGACTCGCGAGCGCGGGGGTG   120
          ||||||||||| |||| || |||||  ||| || |||||| ||||||||||| |||||||
Db    131 TCGGGGACCACCTGCCGCAGCCATGCGCAGCCCGCGGCCCCCAGGAGCTAACGGGCCTG   190

Qy    121 GTTCGGAGGACATGGGACTACAGTCAAGTCAGTCATCAATGGAGATGTGTTTCAGGAGG   180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    191 GTTCTGGAGGACATGGGACTACAGTCAAGTCAGTCATCAATGGAGATGTGTTTCAGGAGG   250

Qy    181 CCGATGCTACGCTGCGATTGCCCAGGCGACGAGGTTGACTCAGGACCCTGAGAGCATCC   240
          ||| |||||| ||||||| || |||||||||||||||||| |||||||||||||||||||
Db    251 CTGATGCTACGCTGCGATTGCCCAGGCTGACAGGCTGACCCAGGAGCCTGAGAGCATCC   310

Qy    241 GCAAGTGGAGAGAGGAGCAGCGCAGAAGAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA  300
          ||| ||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db    311 GCAAGTGGCGAGAGGAGCAGCGAGAGCGACTGCAAGACTCGAAGACTGCTGCATCATCAAGGTCA  370

Qy    301 CCGAACAGGAGTGGCGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA  360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    371 CCGAACAGGAATGGCGCGGGAGAAGGCCAAGACACAGAGAGGCCAAGACACTGGAGGAGTGGAACCAGCGCCAGA  430

Qy    361 GTGAACAGGTTGAGAAGAACAAGATCAACAACACAGGGCATCGGAAGAGAGCGCTTTTTGTGAAAG  420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    431 GTGAACAAGTAGAGAAGAACAAGATCAACAACACCGGGCATCCGGGCCACCCGGGCAGGCTTTTCGTGAAGG  490

Qy    421 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCA  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    491 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTATGTGACTTCA  550

Qy    481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGGCTGCGCTCCGTGCTCAT  535
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Db   551 ACCCCAAGAGCAGCAAGCAGTGCAAAGATGTGTCCCGGCTGCGCTCGGTGCTCAT   605

```
Db
RESULT 6
LOCUS       CR536577              636 bp    mRNA    linear   PRI 17-APR-2005
DEFINITION  Homo sapiens full open reading frame cDNA clone RZPDo834B0322D for
            gene CLTB, clathrin, light polypeptide (Lcb); complete cds, incl.
            stopcodon.
ACCESSION   CR536577
VERSION     CR536577.1  GI:49168637
KEYWORDS    Full ORF shuttle clone, Gateway(TM), complete cds.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 636)
  AUTHORS   Halleck,A., Ebert,L., Mkoundinya,M., Schick,M., Eisenstein,S.,
            Neubert,P., Kstrang,K., Schatten,R., Shen,B., Henze,S., Mar,W.,
            Korn,B., Zuo,D., Hu,Y. and LaBaer,J.
  TITLE     Cloning of human full open reading frames in Gateway(TM) system
            entry vector (pDONR201)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 636)
  AUTHORS   Halleck,A., Ebert,L., Mkoundinya,M., Schick,M., Eisenstein,S.,
            Neubert,P., Kstrang,K., Schatten,R., Shen,B., Henze,S., Mar,W.,
            Korn,B., Zuo,D., Hu,Y. and LaBaer,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUN-2004) RZPD Deutsches Ressourcenzentrum fuer
            Genomforschung GmbH, Im Neuenheimer Feld 580, D-69120 Heidelberg,
            Germany
COMMENT     RZPD; RZPDo834B0322D, ORFNo 3192
            www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=RZPDo834B0322D RZPDLIB;
            Human Full ORF Clones Gateway(TM) - RZPD (kan-resist.) RZPD LIB No.
            834
            www.rzpd.de/cgi-bin/products/showLib.pl.cgi/response?libNo=834
            www.rzpd.de/products/orfclones/
            Contact: Inge Arlart
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH,
            Heubnerweg 6, D-14059 Berlin, Germany
            Tel: +49 30 32639 100
            Fax: +49 30 32639 111
            www.rzpd.de
            This clone is available from RZPD;
            contact RZPD (customer.service@rzpd.de) for further information.
            This CDS clone is a part of a collection of human full ORF clones
            jointly established and verified by the Harvard Institute of
            Proteomics and RZPD.
            This CDS has been cloned incl. stopcodon.
            The CDS has been inserted into pDONR201 via a BP Clonase(TM)
            reaction. Additional sequence has been added in front of the start
            codon: att...AAAAAA GCA GGC TCC ACC (ATG).
            The stopcodon is followed by the 3' att site:
            (stop)GACCCAGCTTTCTT...att Compared to the reference sequence
            NM_001834 (gi32483393) we found AA exchange(s) at position (first
            base of changed triplet): 13(phe->ile)
            Clone distribution: http://www.rzpd.de/products/orfclones/.
FEATURES             Location/Qualifiers
     source          1..636
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="RZPD:RZPDo834B0322D"
                     /db_xref="taxon:9606"
                     /clone="RZPDo834B0322D"
                     /clone_lib="Human Full ORF Clones Gateway(TM) - RZPD"
                     /lab_host="DH10B"
                     /note="Vector: pDONR201, Site_1: attP1; Site_2: attP2"
     gene            1..636
                     /gene="CLTB"
     CDS             1..636
                     /gene="CLTB"
                     /codon_start=1
```

```
            /protein_id="CAG38814.1"
            /db_xref="GI:49168638"
            /db_xref="GOA:P09497"
            /db_xref="UniProt/Swiss-Prot:P09497"
            /translation="MADDIGFPSSSESGAPEAAEEDPAAAFLAQQESEIAGIENDEGF
            GAPAGSHAAPAQPGPTSGAGSEDMGTTVNGDVFQEANGPADGYAAIAQADRLTQEPES
            IRKWREEQRKRLQELDAASKVTEQEWREKAKKDLEEWNQRSEQVEKNKINNRASEEA
            FVKESKEETPGTEWEKVAQLCDFNPKSSKQCKDVSRLRSVLMSLKQTPLSR"
ORIGIN

Query Match      84.4%;    Score 451.8;   DB 8;   Length 636;
Best Local Similarity   90.3%;   Pred. No. 2e-92;
Matches  483;   Conservative   0;   Mismatches   52;   Indels   0;   Gaps   0;

Qy        1 CGGGCCGCCTTCCTGGCCCCAGCAGGAGGAGGAGATTGCTGGCATCGAGAATGACTCGGGTT 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       71 CGGGCCGCCTTCCTGGCCCCAGCAGGAGGAGGAGATTGCTGGCATCGAGAATGACGAGGGCT 130

Qy       61 TCGGGGACACCTGCCGCGCCAGCAGCCAGTGCCTTCGCGAGACCTCGCCGAGCGGCGGGGTG 120
            |||||||||||||||||| ||||| |||||||||||||||||||||||||||||||||||
Db      131 TCGGGGACACCTGCCGCGCCAGCAGCCATGCGGCCCCCGCGCCAGCGGGCCCCACGAGTGGGGCTG 190

Qy      121 GTTCGGAGGACATGGGGACTCAATGGAGATGTGTTCAGGAGGCTAACGGGCTCAT 180
            |||||||||||||||||||||||||| ||||||||||||||||||||||||
Db      191 GTTCGGAGGACATGGGGACCACCAGTCAATGGAGATGTGTTTCAGGAGGCCAACGGGCTCAT 250

Qy      181 CCGATGGCTACGCTGCGATTGCCCAGGCCGGACAGGTTGACTCAGGAGGCTGAGAGCATCC 240
            ||||||||||||| ||| ||||||||||||||||||||||||||||||||||||||||
Db      251 CTGATGGCTACGCACGGCCATTGCCCAGGCTGACAGGGCTGACAGGCCTGAGAGCATCC 310

Qy      241 GCAAGTGGAGACGAGAGGAGCAGAAGCTCGCAGGAGTTGGATGCTGCCTCGAAGGTGA 300
            |||||||||||||||| |||| |||||||||||||||||||||| |||||||||||||
Db      311 GCAAGTGGAGCAGAGGAGGAGCAGAAACGCGTCAAGAGCTCGGAAGAGTGGGGGCTG 370

Qy      301 CCGAACAGGAGTGGCGGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 360
            |||||||||||| |||||||||||||||||||||||||| ||||||||||||||||
Db      371 CGGAACAGGAGGAATGGCGGGGAGAAGGCCAAGAAGAGGAGGCCAAGAAAGACCTGGAACAGCGCCAGA 430

Qy      361 GTGAACAGGTTGAGAGAAGAACAAGATCAACAACAGGGCATCGGAAGAGAGGCTTTTGTGAAAG 420
            ||||||||| |||||| ||||||||||||||||||||||||||||||| ||||||||
Db      431 GTGAACAAGTAGAGAGAAGAACAAGATCAACAACACCGGGCATCGGGAGAGAGGCTTCGTGAAGG 490

Qy      421 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGCCCAGCTGTGTGACTTCA 480
            ||||||||||||||||||||||||||||||||||||||||||||||| |||| |||||
Db      491 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTATGTGACTTCA 550

Qy      481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTGCGTTCGGTGCTCAT 535
            |||||| |||||||||||||| ||||||||||||||||||||||||||||||||
Db      551 ACCCCAAGAGCAGCAAGCACTGCAAAGATGTGTCCCGCCTGCGCTCGGTGCTCAT 605
```

```
RESULT 7
AY889907
LOCUS        AY889907               636 bp    mRNA     linear    SYN 29-MAR-2005
DEFINITION   Synthetic construct Homo sapiens clone FLH025393.01X clathrin light
             polypeptide (CLTB) mRNA, complete cds.
ACCESSION    AY889907
VERSION      AY889907.1  GI:60655268
KEYWORDS     Human ORF Project.
SOURCE       synthetic construct
  ORGANISM   synthetic construct
             other sequences; artificial sequences.
REFERENCE    1  (bases 1 to 636)
  AUTHORS    Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
             Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,B.,
             Williamson,J. and LaBaer,J.
  TITLE      Cloning of human full-length CDS in Creator (TM) recombinational
             vector system
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 636)
  AUTHORS    Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
             Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,B.,
             Williamson,J. and LaBaer,J.
  TITLE      Direct Submission
  JOURNAL    Submitted (04-JAN-2005) Biological Chemistry and Molecular
             Pharmacology, Harvard Institute of Proteomics, 320 Charles St.,
             Cambridge, MA 02141, USA
COMMENT      This CDS clone is a part of a collection of human full-length
             expression clones generated by Harvard Institute of Proteomics.
             This ORF clone has been cloned with normalized stop-codon. The CDS
             has been directionally cloned using BD In-Fusion(TM) cloning system
             between the SalI and HindIII sites of the pDNR-Dual vector.
             Additional sequences in the clone:  'ACC' after SalI site and
             before 'ATG' to provide Kozak consensus sequence. Each clone is
             clonally isolated and full-length sequence-verified.
FEATURES             Location/Qualifiers
     source          1..636
                     /organism="synthetic construct"
                     /mol_type="mRNA"
                     /db_xref="taxon:32630"
                     /clone="FLH025393.01X"
                     /lab_host="Escherichia coli DH5alpha T1 resistant"
                     /note="derived from MGC template"
     gene            1..636
                     /gene="CLTB"
     CDS             1..636
                     /gene="CLTB"
                     /note="Lcb"
                     /codon_start=1
                     /transl_table=11
                     /product="clathrin light polypeptide"
                     /protein_id="AAX32698.1"
                     /db_xref="GI:60655269"
                     /translation="MADDFGFPSSSESGAPEAAEEDPAAAFLAQQESEIAGIENDEGF
                     GAPAGSHAAPAQPGPTSGAGSEDMGTTVNGDVFQEANGPADGYAAIAQADRLTQEPES
                     IRKWREEQRKRLQELDAASKVTEQEWREKAKKDLEEWNQRSEQVEKNKINNRASEEA
                     FVKESKEETPGTEWEKVAQLCDFNPKSSKQCKDVSRLRSVLMSLKQTPLSR"
ORIGIN

Query Match      84.4%;    Score 451.8;   DB 11;   Length 636;
Best Local Similarity   90.3%;   Pred. No. 2e-92;
Matches  483;   Conservative   0;   Mismatches   52;   Indels   0;   Gaps   0;

Qy        1 CGGGCCGCCTTCCTGCCCCAGCAGGAGGAGGAGATTGCTGGCATCGAGAATGACTCGGGTT 60
            |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       71 CTGATGCTACGCAGCCATTGCCCAGGCTGACAGGCCTGACCCAGGAGGCCTGAGAGCATCC 310

Qy        1 CGGGCCGCCTTCCTGCCCCAGCAGGAGGAGGAGATTGCTGGCATCGAGAATGACGAGGGCT 130
Db       71

Qy       61 TCGGGGCACCTGCCGCGCCAGTGCCTTCGCGAGACCTCGCCGAGCGGCGGGGTG 120
            |||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db      131 TCGGGGCACCTGCCGCGCCAGCATGCGGCCCCCGCGGGCCCCACGAGTGGGGCTG 190

Qy      121 GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTCAGGAGGCTAACGGGCCTG 180
            |||||||||||||||||||||| |||||||||||||||||| ||||||||||||
Db      191 GTTCTGAGGACATGGGGACCACCAGTCAATGGAGATGTGTTTCAGGAGGCCAACGGGCTCTG 250

Qy      181 CCGATGGCTACGCTGCGATTGCCCAGGCCGGACAGGTTGCCCAGGAGGCCTGAGAGCATCC 240
            ||||| ||||||||||||||||||||||||||| |||||||||||||||||||||||
Db      251 CTGATGCTACGCAGCCATTGCCCAGGCTGACAGGCCTGACCCAGGAGGCCTGAGAGCATCC 310

Qy      241 GCAAGTGGAGAGAGGAGCAGAAGCTCGCAGGAGATTGCAGGCATAGAGAACGACGAGGGCT 300
            |||||||||||||||||| ||||||||||||||||||||||||||| ||||||||||||
Db      311 GCAAGTGGAGCAGAGAGAGCAGAAACGCGTCAAGAGCTCGAAGAGTGCATCTAAGGTCA 370

Qy      301 CCGAACAGGAGTGGCGGGGAGAAGGCCAAAAAAAGACCTGGAGGAGTGAACCAGCGCCAAA 360
            |||||||||||||||||||||||||||||||||||| |||||||||| ||||||||||
Db      371 CGGAACAGGAGGATGGCGGGGAGAAGGCCAAGAAGAGGAGGCCAAGAACAGCGCCAGA 430

Qy      361 GTGAACAGGTTGAGAGAGAACAAGATCAACAACAGGGCATCGGAAGAGAGGCTTTTGTGAAAG 420
            ||||||||| ||||||||||||||||||||||||||||||||||||||| ||||||||
Db      431 GTGAACAAGTAGAGAGAAGAACAAGATCAACAACACCGGGCATCGGGAGAGAGGCTTCGTGAAGG 490

Qy      421 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCA 480
            ||||||||||||||||||||||||||||||||||||||||| ||||||||| ||||||||
Db      491 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTATGTGACTTCA 550
```

```
Qy   61  TCGGGCCACCTGCCGCCAGCCAGGTGGCCTCTGCGCCTCGGCGACCTCGGAGGCGGGGGTG  120
         ||||||||||||||||||  |||||||||||  ||  | ||  |||||||||||
Db  131  TCGGGGCACCTGCCGCCAGCCAGCCATGGGGCCGCGGCGCCCCGGCGGCCCCACGAGTCGGGGCTG  190

Qy  121  GTTCGGAGGACATGGGGACTACAGTCAATGATGTGTTCAGGAGGCGTAACGGGCCTG  180
         |||| ||||| ||||||  |  | || |||||||||||||||||||||
Db  191  GTTCTGAGGACATGGGGACCACAGTCAATGATGTGTTCAGGAGGCCAACGGTCCTG  250

Qy  181  CCGATGGCTACGCTGCGATTGCCCAGGCGACAGGTTGACTCAGGAGCTCAGGAGCATCC  240
         ||||||||| |||||| ||||||||||||  ||  ||||||||||||||||
Db  251  CTGATGGCTACGACGCCATTGCCCAGGCTGACCCAGGCGTGGATGCCGCTCGAAGGTGA  310

Qy  241  GCAAGTGGAGAGGAGCAGAGAAGCTGCAGGAGGTGGATGCCGCTCGAAGGTGA  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  311  GCAAGTGGAGAGGAGCAGAGAGAAACGGCTGGAAGACTGGGAGAGTGCTGAGGAGTGCTCATCCTAAGGTCA  370

Qy  301  CCGAACAGGAGTGGCGGGAGGAGGCCAAAAAGACCTTGGAGAGTGGAACCAGCCGCCAAA  360
         ||||  |||||  |||||||  ||||||||||||||||||||||||||
Db  371  CGGAACAGGAATGCGCGGGAAGGCCAAGAAGACCTTGGAGGACACCAGCACCGCCCAGA  430

Qy  361  GTGAACAGTTGAGAAGAACAAGATCAACAACAGGGCATCGGAAGAGGCTTTTGTGAAAG  420
         ||||||| ||||||||||||||| |||||||||||  | |||||||||||
Db  431  GTGAACAAGTAGAGAAGAACAAGATCAACAACCGGGCATCCGGCATCCGGAGGAGGCTTTCGTGAAGG  490

Qy  421  AATCCAAGGAGGACCCCAGGACACAGAGTGGGAGAAGTGGGCCCAGCTGTGTGACTTCA  480
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  491  AATCCAAGGAGGACCCCAGGACACAGAGTGGGAGAAGTGGGCCCAGCTATGTGACTTCA  550

Qy  481  ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTCGGCTCGGTGCTCAT  535
         ||||  |||||||||||||||||||||||||||||||||||||||||||||||
Db  551  ACCCCAAGAGCAGCAAGCAGTGCAAAGATGTGTCCCGCCTCGGCTCGGTGCTCAT  605
```

```
RESULT 8
AY892384
LOCUS        AY892384                636 bp    mRNA    linear   SYN 29-MAR-2005
DEFINITION   Synthetic construct Homo sapiens clone FLH025388.01L clathrin light
             polypeptide (CLTB) mRNA, partial cds.
ACCESSION    AY892384
VERSION      AY892384.1  GI:60653212
KEYWORDS     Human ORF project.
SOURCE       synthetic construct
  ORGANISM   synthetic construct
             other sequences; artificial sequences.
REFERENCE    1  (bases 1 to 636)
  AUTHORS    Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
             Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
             Williamson,J. and LaBaer,J.
  TITLE      Cloning of human full-length CDS in Creator (TM) recombinational
             vector system
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 636)
  AUTHORS    Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
             Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
             Williamson,J. and LaBaer,J.
  TITLE      Direct Submission
  JOURNAL    Submitted (05-JAN-2005) Biological Chemistry and Molecular
             Pharmacology, Harvard Institute of Proteomics, 320 Charles St.,
             Cambridge, MA 02141, USA
COMMENT      This CDS clone is a part of a collection of human full-length
             expression clones generated by Harvard Institute of Proteomics.
             This ORF clone has been cloned without stop-codon (to allow fusion
             with C-terminal tag). The CDS has been directionally cloned using
             BD In-Fusion(TM) cloning system between the SalI and HindIII sites
             of the pDNR-Dual vector. Additional sequences in the clone: 'ACC'
             after SalI site and before 'ATG' to provide Kozak consensus
             sequence; 'GG' after last codon and before HindIII site to maintain
             reading frame. Each clone is clonally isolated and full-length
             sequence-verified.
FEATURES             Location/Qualifiers
     source          1..636
                     /organism="synthetic construct"
                     /mol_type="mRNA"
                     /db_xref="taxon:32630"
                     /clone="FLH025388.01L"
                     /lab_host="Escherichia coli DH5alpha T1 resistant"
                     /note="derived from MGC template"
     gene            1..>636
                     /gene="CLTB"
     CDS             1..>636
                     /gene="CLTB"
                     /note="Lcb"
                     /codon_start=1
                     /transl_table=11
                     /product="clathrin light polypeptide"
                     /protein_id="AAX29301.1"
                     /db_xref="GI:60653213"
                     /translation="MADDFGFFSSSESGAPEAAEEDPAAAFLAQQESBIAGIENDEGF
                     GAPAGSHAAPAQPGTGAGSBDMGTTVNGDVFQEANGPADGYAIAQADRLTQEPES
                     IRKWREEQRKRLQELDAASKVTEQEWREKAKKDLEEWNQRSEQVEKNKINNRASEEA
                     FVKESKEETPGTENEKVAQLCDFNPKSSKQCKDVSRLRSVLMSLKQTPLSRL"
ORIGIN
Query Match       84.4%;   Score 451.8;  DB 11;  Length 636;
Best Local Similarity   90.3%;  Pred. No. 2e-92;
Matches  483;   Conservative  0;  Mismatches  52;  Indels  0;  Gaps  0;

Qy    1  CGGCCGCCTTCCTGGCCCAGGAGAGCGGAGATTGCTGGCATCGAGAATGACTCGGGTT  60
         ||||||||||||||||||||||||||||||  |  ||||||||||   | |||
Db   71  CGGCCCGGCCTTCCTGGCCCAGCAGGAGAGCGGAGATTGCAGGCATAGAGAACGACGAGGGCT  130
```

```
Qy  481  ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTCGGCTCGGTGCTCAT  535
         ||||  |||||||||||||||||||||||||||||||||||||||||||||||
Db  551  ACCCCAAGAGCAGCAAGCAGTGCAAAGATGTGTCCCGCCTCGGCTCGGTGCTCAT  605

RESULT 9
AY892385
LOCUS        AY892385                636 bp    mRNA    linear   SYN 29-MAR-2005
DEFINITION   Synthetic construct Homo sapiens clone FLH025389.01L clathrin light
             polypeptide (CLTB) mRNA, partial cds.
ACCESSION    AY892385
VERSION      AY892385.1  GI:60653214
KEYWORDS     Human ORF project.
SOURCE       synthetic construct
  ORGANISM   synthetic construct
             other sequences; artificial sequences.
REFERENCE    1  (bases 1 to 636)
  AUTHORS    Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
             Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
             Williamson,J. and LaBaer,J.
  TITLE      Cloning of human full-length CDS in Creator (TM) recombinational
             vector system
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 636)
  AUTHORS    Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
             Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
             Williamson,J. and LaBaer,J.
  TITLE      Direct Submission
  JOURNAL    Submitted (05-JAN-2005) Biological Chemistry and Molecular
             Pharmacology, Harvard Institute of Proteomics, 320 Charles St.,
             Cambridge, MA 02141, USA
COMMENT      This CDS clone is a part of a collection of human full-length
             expression clones generated by Harvard Institute of Proteomics.
             This ORF clone has been cloned without stop-codon (to allow fusion
             with C-terminal tag). The CDS has been directionally cloned using
             BD In-Fusion(TM) cloning system between the SalI and HindIII sites
             of the pDNR-Dual vector. Additional sequences in the clone: 'ACC'
             after SalI site and before 'ATG' to provide Kozak consensus
             sequence; 'GG' after last codon and before HindIII site to maintain
             reading frame. Each clone is clonally isolated and full-length
             sequence-verified.
FEATURES             Location/Qualifiers
     source          1..636
                     /organism="synthetic construct"
                     /mol_type="mRNA"
```

```
gene             /db_xref="taxon:32630"
                 /clone="PLH025389.01L"
                 /lab_host="Escherichia coli DH5alpha T1 resistant"
                 /note="derived from MGC template"
                 1..>636
                 /gene="CLTB"
CDS              1..>636
                 /gene="CLTB"
                 /note="Lcb"
                 /codon_start=1
                 /transl_table=11
                 /product="clathrin light polypeptide"
                 /protein_id="AAX29302.1"
                 /db_xref="GI:6053215"
                 /translation="MADDFGFFSSSESGAPEAAEDPAAAFLAQQESEIAGIENDEGF
                 GAPAGSHAAPAQPGPTSGAGSEDMGTTVNGDVFQEANGPADGYAAIAQADRLTQEPES
                 IRKWREEQRKRLQELDAASKVTEQEWREKAKKDLEEWNQRQSEQVEKNKINNRASEEA
                 FVEKSKEETPGTEWEKVAQLCDFNPKSSKQCKDVSRLRSVLMSLKQTPLSRL"
ORIGIN

Query Match     84.4%;  Score 451.8;  DB 11;  Length 636;
Best Local Similarity   90.3%;  Pred. No. 2e-92;
Matches  483;  Conservative  0;  Mismatches  52;  Indels  0;  Gaps  0;

Qy     1  CGGCCGCCTTCCTGGCCCAGCAGGAGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT  60
          ||||||||||||||||||||||||||||||| ||| ||| |||||||||||||||||||
Db    71  CGGCCGCCTTCCTGGCCCAGCAGGAGAGCGAGATTGCAGGCATAGAGAACGACGAGGGCT 130

Qy    61  TCGGGGCACCTGCCCGCCGCCCAGTGGCCTCTGCGCCCGGGACTCGCCGAGCGCGGGGTG 120
          ||||||||||||||| |||||||||||||| || ||||||||||| |||||||||||||
Db   131  TCGGGGCACCTGCCCGGCCGCCCAGTGGCCCAGTGGGGCCCCACGGGCCTAACGGGCCTG 190

Qy   121  GTTCGGAGGACATGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACGGGCCTG   180
          |||| |||||||||||||| ||||||||||||||||||||||||||| ||||||||||
Db   191  GTTCTGAGGACATGGGACCACCAGTCAATGAGATGTGTTTCAGGAGGCCAACGGTCCTG  250

Qy   181  CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGTTGACTCAGGAGCTGAGAGCATCC   240
          | |||||||||| | |||||| ||||||||||| |||||||| ||||||||||||||
Db   251  CTGATGGCTACGCAGCCATTGCCCAGGCTGACACAGGCTGACCCAGGAGCTGAGAGCATCC 310

Qy   241  GCAAGTGGAGAGAGGAGCAGAGAGAAAGGCTGCAGGAGTTGGATGCGTCGAAGTTGA    300
          ||||||||| ||||||||||||||| |||||||||||| ||||||||| | |||||
Db   311  GCAAGTGGCGCGAGGAGCAGAGAGGAAACGGCTGCAGGAGCTGGATGCGTCATCTAAGGTCA 370

Qy   301  CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 360
          |||| ||||||||||| ||||||| |||||||||||||||||||| ||||||||||||||
Db   371  CGGAACAGGAATGGCGGGAGAAGGCCAAGAAAGACCTGGAGGAGTGGAACCAGCGCCAGA  430

Qy   361  GTGAACAGGTTGAGAAGAGACAAGATCAACAACAGGGCATCAACAACCGGGCATCAACTTC 420
          |||||||||||||| ||||| |||||| |||||||||||| ||||||||||||  ||||
Db   431  GTGAACAGGTTGAGAAGAGACAAGATCAACAACAGGGCCTGACAGGCTGACCCAGGAGG  490

Qy   421  AATCCAAGGAGGAGACCCCAGGCACAGAGTGGCCCAGCTGTGTGACTTCA         480
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   491  AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGCCCAGCTGTGTGACTTCA        550

Qy   481  ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTGCGTCCGGCTCAT 535
          |||| ||||||||||||||||||| |||||||||||||||||||||||||||
Db   551  ACCCCAAGAGCAGCAAGCAGTGCAAAGATGTGTCCCGCCTGCGTCCGGCTCAT 605


RESULT 10
BT007759

LOCUS       BT007759                636 bp    mRNA     linear   SYN 13-MAY-2003
DEFINITION  Synthetic construct Homo sapiens clathrin, light polypeptide (Lcb)
            mRNA, partial cds.
ACCESSION   BT007759
VERSION     BT007759.1  GI:30584356
KEYWORDS    FLI_CDNA.
SOURCE      synthetic construct
ORGANISM    synthetic construct
            other sequences; artificial sequences.
REFERENCE   1 (bases 1 to 636)
```

```
AUTHORS     Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
TITLE       Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 636)
AUTHORS     Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
TITLE       Direct Submission
JOURNAL     Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
            Circle, Palo Alto, CA 94303, USA
COMMENT     This CDS clone is a part of a collection of human full length
            expression clones generated by BD Biosciences Clontech and the
            Harvard Institute of Proteomics. Each CDS has been cloned in two
            forms: with and without stop-codon (to allow fusion with C-terminal
            tag). The CDS has been directionally cloned using BD In-Fusion(TM)
            cloning system between the SalI and HindIII sites of the pDNR-DUAL
            vector. Additional sequences in the clone: 'ACC' after SalI site
            and before 'ATG' to provide Kozak consensus sequence; 'GG' after
            last codon and before HindIII site to maintain reading frame.
            Clone distribution: http://bioinfo.clontech.com/orfclones
FEATURES             Location/Qualifiers
     source          1..636
                     /organism="synthetic construct"
                     /mol_type="mRNA"
                     /db_xref="taxon:32630"
                     /clone="GH00332L1.0"
                     /clone_lib="BD Creator(TM) CDS Library derived from MGC
                     collection"
                     /lab_host="DH5alpha T1 resistant"
                     /note="Vector: pDNR-Dual"
     CDS             1..>636
                     /note="Mutations: 635:Stop->Leu"
                     /codon_start=1
                     /transl_table=11
                     /product="Homo sapiens clathrin, light polypeptide (Lcb)"
                     /protein_id="AAP36427.1"
                     /db_xref="GI:30584357"
                     /translation="MADDFGFFSSSESGAPEAAEDPAAAFLAQQESEIAGIENDEGF
                     GAPAGSHAAPAQPGPTSGAGSEDMGTTVNGDVFQEANGPADGYAAIAQADRLTQEPES
                     IRKWREEQRKRLQELDAASKVTEQEWREKAKKDLEEWNQRQSEQVEKNKINNRASEEA
                     FVEKSKEETPGTEWEKVAQLCDFNPKSSKQCKDVSRLRSVLMSLKQTPLSRL"
ORIGIN

Query Match     84.4%;  Score 451.8;  DB 11;  Length 636;
Best Local Similarity   90.3%;  Pred. No. 2e-92;
Matches  483;  Conservative  0;  Mismatches  52;  Indels  0;  Gaps  0;

Qy     1  CGGCCGCCTTCCTGGCCCAGCAGGAGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT  60
          ||||||||||||||||||||||||||||||| ||| ||| |||||||||||||||||||
Db    71  CGGCCGCCTTCCTGGCCCAGCAGGAGAGCGAGATTGCAGGCATAGAGAACGACGAGGGCT 130

Qy    61  TCGGGGCACCTGCCCGCCGCCCAGTGGCCTCTGCGCCCGGGACTCGCCGAGCGCGGGGTG 120
          ||||||||||||||| |||||||||||||| || |||||||||||| ||||||||||||
Db   131  TCGGGGCACCTGCCCGGCCGCCCAGTGGCCCAGTGGGGCCCCACGGGCCTAACGGGCCTG 190

Qy   121  GTTCGGAGGACATGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACGGGCCTG   180
          |||| |||||||||||||| ||||||||||||||||||||||||||| ||||| ||||
Db   191  GTTCTGAGGACATGGGACCACCAGTCAATGAGATGTGTTTCAGGAGGCCAACGGTCCTG  250

Qy   181  CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGTTGACTCAGGAGCTGAGAGCATCC   240
          | |||||||||| | |||||| ||||||||||| |||||||| ||||||||||||||
Db   251  CTGATGGCTACGCAGCCATTGCCCAGGCTGACACAGGCTGACCCAGGAGCTGAGAGCATCC 310

Qy   241  GCAAGTGGAGAGAGGAGCAGAGAGAAAGGCTGCAGGAGTTGGATGCGTCGAAGTTGA    300
          ||||||||| ||||||||||||||| |||||||||||| ||||||||| | ||||||
Db   311  GCAAGTGGCGCGAGGAGCAGAGAGGAAACGGCTGCAGGAGCTGGATGCGTCATCTAAGGTCA 370

Qy   301  CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 360
          |||| ||||||||||| ||||||| |||||||||||||||||||| ||||||||||||||
Db   371  CGGAACAGGAATGGCGGGAGAAGGCCAAGAAAGACCTGGAGGAGTGGAACCAGCGCCAGA  430

Qy   361  GTGAACAGGTTGAGAAGAGACAAGATCAACAACAGGGCATCAACAACCGGGCATCAACTTC 420
          |||||||||||||| ||||| |||||| |||||||||||| ||||||||||| ||||||
Db   431  GTGAACAGGTTGAGAAGAGACAAGATCAACAACAGGGCCTGACAGGCTGACCCAGGAGG  490

Qy   421  AATCCAAGGAGGAGACCCCAGGCACAGAGTGGCCCAGCTGTGTGACTTCA         480
          ||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   491  AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGCCCAGCTGTGTGACTTCA        550

Qy   481  ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTGCGTCCGGCTCAT 535
          |||| ||||||||||||||| || |||||||||||||||||||||||||||||
Db   551  ACCCCAAGAGCAGCAAGCAGTGCAAAGATGTGTCCCGCCTGCGTCCGGCTCAT 605
```

```
Qy  181 CCGATGGCTACGCTGCGATTGCCCAGGCGGCGACAGGTTGACTCAGGAGCCTGAGAGCATCC 240
        ||||||||||  ||| |||||||||||||||||||||||||||||  |||||||||||||||
Db  390 CTGATGGCTACGCCCAGGCCATTGCCCAGGCTGACAGGCTGACCCAGGAGCCTGAGAGCATCC 449

Qy  241 GCAAGTGGAGAGAGGAGCAGAGAAAAAGGCTGCAGAGTTGATGCTGCTGGCCTCGAAGGTGA 300
        ||||||||  |||||||||| ||||| ||||||||  |||| ||||||| || |||||||| |
Db  450 GCAAGTGCGCGAGGAGCAGCGGAAAAGGCTGCAAGAGCTGGATGCTGCAAGACTCTAAGGTCA 509

Qy  301 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTTGAGGAGTGGAACCAGCGCCAAA 360
        | |||||||| |||||||||||||||||| || ||||| |||||||||||||||||| |
Db  510 CGGAACAGGAATGGCGGGAGAAGGCCAAGAAGGACCTGGAGGAGTGGAACCAGCGCCAGA 569

Qy  361 GTGAACAGGTTGAGAAGAACAAGATCAACAACACAGGGCATCGGAAGAGGCTTTTGTGAAAG 420
        ||||||| || ||||||||||||||||||||||||| ||||||| ||||||||| |||||| |
Db  570 GTGAACAAGTAGAGAAGAACAAGATCAACAACACCGGGCATCGGAGAGGGCTTTCGTGAAGG 629

Qy  421 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGCCCAGCTGTGTGACTTCA 480
        |||||||||||||||||||||||||||||||||  ||||| |||||||||||
Db  630 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGCGTGCCCAGCTATGTGACTTCAT 689

Qy  481 ACCCTAAGAGCAGCAAGCAAGTGTAAAGACGTGTTCCCGCCTCGGTGCTCAT 535
        |||| |||||||||||||||||||  ||||  ||||| ||||||||||| ||||
Db  690 ACCCCAAGAGCAGCAAGCAAGTGTCAAAGATGTGTCCCGCCTCGGCTCGGTGCTCAT 744
```

RESULT 12
BC006457
LOCUS   BC006457   1099 bp   mRNA   linear   PRI 08-MAR-2005
DEFINITION Homo sapiens clathrin, light polypeptide (Lcb), transcript variant
            nonbrain, mRNA (cDNA clone MGC:1776 IMAGE:3538261), complete cds.
ACCESSION   BC006457
VERSION   BC006457.1 GI:13623662
KEYWORDS   MGC.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 1099)
AUTHORS   Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD,
            Collins FS, Wagner L, Shenmen CM, Schuler GD, Altschul SF, Zeeberg
            B, Buetow KH, Schaefer CF, Bhat NK, Hopkins RF, Jordan H, Moore T,
            Max SI, Wang J, Hsieh F, Diatchenko L, Marusina K, Farmer AA, Rubin
            GM, Hong L, Stapleton M, Soares MB, Bonaldo MF, Casavant TL,
            Scheetz TE, Brownstein MJ, Usdin TB, Toshiyuki S, Carninci P,
            Prange C, Raha SS, Loquellano NA, Peters GJ, Abramson RD, Mullahy
            SJ, Bosak SA, McEwan PJ, McKernan KJ, Malek JA, Gunaratne PH,
            Richards S, Worley KC, Hale S, Garcia AM, Gay LJ, Hulyk SW,
            Villalon DK, Muzny DM, Sodergren EJ, Lu X, Gibbs RA, Fahey J,
            Helton E, Ketteman M, Madan A, Rodrigues S, Sanchez A, Whiting M,
            Madan A, Young AC, Shevchenko Y, Bouffard GG, Blakesley RW,
            Touchman JW, Green ED, Dickson MC, Rodriguez AC, Grimwood J,
            Schmutz J, Myers RM, Butterfield YS, Krzywinski MI, Skalska U,
            Smailus DE, Schnerch A, Schein JE, Jones SJ and Marra MA.
CONSRTM   Mammalian Gene Collection Program Team
TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED   12477932
REFERENCE   2 (bases 1 to 1099)
AUTHORS   Director MGC Project.
TITLE   Direct Submission
JOURNAL   Submitted (09-APR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
COMMENT   Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
```

```
Qy  361 GTGAACAGGTTGAGAAGAACAAGATCAACAACACAGGGCATCGGAAGAGGCTTTTGTGAAAG 420
        ||||||| ||  ||||||||||||||||||||||  |||||||| || ||||||| |||||| |
Db  431 GTGAACAAGTAGAGAAGAACAAGAACAGATCAACAACACCGGCATCGGAGAGGGCTTTCGTGAAGG 490

Qy  421 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGCCAGCTGTGTGACTTCA 480
        ||||||||||||||||||||||||||||||||| |||||||||| |||||
Db  491 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGCGTGCCCAGCTATGTGACTTCAT 550

Qy  481 ACCCTAAGAGCAGCAAGCAAGTGTAAAGACGTGTTCCCGCCTCGGTGCTCAT 535
        |||| |||||||||||||||||||  |||  |||||| |||||||||| ||||
Db  551 ACCCCAAGAGCAGCAAGCAAGTGTCAAAGATGTGTCCCGCCTCGGCTCGGTGCTCAT 605
```

RESULT 11
HUMLCTHB
LOCUS   HUMLCTHB   1051 bp   mRNA   linear   PRI 07-JAN-1995
DEFINITION Human lymphocyte clathrin light-chain B mRNA, complete cds.
ACCESSION   M20470 J04174
VERSION   M20470.1 GI:187056
KEYWORDS   clathrin; clathrin light chain b.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 1051)
AUTHORS   Jackson,A.P. and Parham,P.
TITLE   Structure of human clathrin light chains. Conservation of light
            chain polymorphism in three mammalian species
JOURNAL   J. Biol. Chem. 263 (32), 16688-16695 (1988)
PUBMED   3267234
COMMENT   Original source text: Human lymphocyte (cell line T7527), cDNA to
            mRNA.
            Draft entry and computer-readable sequence [1] kindly submitted by
            A.Jackson, 02-SEPT-1988.
FEATURES   Location/Qualifiers
     source   1..1051
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /map="Unassigned"
     gene   1..1051
                     /gene="CLTB"
     CDS   140..775
                     /gene="CLTB"
                     /note="clathrin light-chain B"
                     /codon_start=1
                     /protein_id="AAA59506.1"
                     /db_xref="GI:307119"
                     /db_xref="GDB:G00-125-861"
                     /translation="MADDFGFFSSSESGAPEAAEBDPAAAFLAQQESBIAGIENDEGF
                     GAPAGSHAAPAQPGPTSGAGSEDMGTTVNGDVFQEANGPADGYAAIAQADRLTQEPES
                     IRKWREEQRKRLQELDAASKVTEQEWREKAKKDLEEWNQRQSEQVEKNINNRASEEA
                     FVKESKEETPGTEWEKVAQLCDFNPKSSKQCKDVSRLRSVLMSLKQTPLSR"
ORIGIN   33 bp upstream of HpaII site.
```

Query Match   84.4%; Score 451.8; DB 8; Length 1051;
Best Local Similarity   90.3%; Pred. No. 2.1e-92;
Matches   483; Conservative   0; Mismatches   52; Indels   0; Gaps   0;

```
Qy    1 CGGCGGCCTTCCTGGCCCAGCAGGAGGCGAGATTGCTGGCATCGAGAATGACTCGGGTT 60
        ||||||||||||||||||  ||||||||||||||||| ||  ||||  |||||||| ||
Db  210 CGGCCGCCTTCCTGGCCCCAGCAGGAGGCGAGATTGCAGGCCATAGAGGCATAGAACGACGAGGGCT 269

Qy   61 TCGGGGACACCTGCCGCCCAGGTGGCCCTTCGCGCAGCCCGCGGCCCCGCAGCGGGGGTG 120
        ||||| |||||||||||| ||||| ||  || |||||||| ||||||||||||||| |||
Db  270 TCGGGGACACCTGCCGCCGCCAGCCATGCGGCCCGCGGCCCCGCAGCGGCCCCCAGCGAGTGGGGCTG 329

Qy  121 GTTCGGAGGACATGGGGACTACAGTCAGTCAATGAGATGGTGTTTCAGGAGGCTAACGGCCTG 180
        ||||| ||||||||||||| ||||||||| |  |||||| ||||||||||||    |||||||
Db  330 GTTCTGAGGACATGGGGACCACAGTCAATGGAGATGTGTTTCAGGAGGATGTGTTTCAGGAGGCCAACGGCCGGTCCTG 389
```

```
Qy    361  GTGAACAGGTTGAGGAAGAACAAGATCAACAACAGGGCATCGGAAGAGGCTTTGTGAAAG  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    560  GTGAACAAGTAGAGGAAGAACAAGATCAACAACCGGGCATCCGAGGAGGCTTTCGTGAAGG  619

Qy    421  AATCCAAGGAGGAGACCCCAGGCACCAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCA  480
           |||||||||||||||||||| |||||||| |||||||||||||||||||||||||||||||
Db    620  AATCCAAGGAGGAGACCCCAGGCACCAGAGTGGGAGAAGGTGGCCCAGCCATGTGACTTCA  679

Qy    481  ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCTCCGGTGCTCAT  535
           ||||||||||||||||||||   ||||||||||||||||||||||||||
Db    680  ACCCCAAGAGCAGCAAGCAGTGCAAAGATGTGTCCCGCCTCCGGTGCTGCTCAT  734
```

RESULT 13

```
LOCUS       BC006332
DEFINITION  Homo sapiens clathrin, light polypeptide (Lcb), transcript variant
            nonbrain, mRNA (cDNA clone MGC:12930 IMAGE:4299637), complete cds.
ACCESSION   BC006332
VERSION     BC006332.2  GI:33991210
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1164)
AUTHORS     Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD,
            Collins FS, Wagner L, Shenmen CM, Schuler GD, Altschul SF, Zeeberg
            B, Buetow KH, Schaefer CF, Bhat NK, Hopkins RF, Jordan H, Moore T,
            Max SI, Wang J, Hsieh F, Diatchenko L, Marusina K, Farmer AA, Rubin
            GM, Hong L, Stapleton M, Soares MB, Bonaldo MF, Casavant TL,
            Scheetz TE, Brownstein MJ, Usdin TB, Toshiyuki S, Carninci P,
            Prange C, Raha SS, Loquellano NA, Peters GJ, Abramson RD, Mullahy
            SJ, Bosak SA, McEwan PJ, McKernan KJ, Malek JA, Gunaratne PH,
            Richards S, Worley KC, Hale S, Garcia AM, Gay LJ, Hulyk SW,
            Villalon DK, Muzny DM, Sodergren EJ, Lu X, Gibbs RA, Fahey J,
            Helton E, Ketteman M, Madan A, Rodrigues S, Sanchez A, Whiting M,
            Madan A, Young AC, Green ED, Dickson MC, Rodriguez AC, Grimwood J,
            Schmutz J, Myers RM, Butterfield YS, Krzywinski MI, Skalska U,
            Smailus DE, Schnerch A, Schein JE, Jones SJ and Marra MA.
CONSRTM     Mammalian Gene Collection Program Team
TITLE       Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED      12477932
REFERENCE   2  (bases 1 to 1164)
AUTHORS     Director MGC Project.
TITLE       Direct Submission
JOURNAL     Submitted (09-APR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
            On Aug 20, 2003 this sequence version replaced gi:13623458.
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
```

BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,
Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth
Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,
Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prabhu,
Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska,
Duane Smailus, Jeff Stott, Miranda Tsai, George Yang, Jacquie
Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 8 Row: j Column: 21
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 32483393.

```
FEATURES             Location/Qualifiers
source          1..1099
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="MGC:1776 IMAGE:3538261"
                /tissue_type="Lung, small cell carcinoma"
                /clone_lib="NIH_MGC_7"
                /lab_host="DH10B-R"
                /note="Vector: pOTB7"
gene            1..1099
                /gene="CLTB"
                /note="synonym: LCB"
                /db_xref="GeneID:1212"
                /db_xref="MIM:118970"
CDS             130..765
                /gene="CLTB"
                /codon_start=1
                /product="clathrin, light polypeptide, isoform a"
                /protein_id="AAH06457.1"
                /db_xref="GI:13623663"
                /db_xref="GeneID:1212"
                /db_xref="MIM:118970"
                /translation="MADDFGFFSSSESGAPEAAEBDPAAAFLAQQESEIAGIENDEGF
                GAPAGSHAAPAQPGPTSGAGSEDMGTTVNGDVFQEANGPADGYAAIAQADRLTQEPES
                IRKWREEQRKLQELDAASKVTEQEWREKAKKDLEEWNQRQSEQVEKNKINNRASERA
                FVKESKEETPGTEWEKVAQLCDFNPKSSKQCKDVSRLRSVLMSLKQTPLSR"
ORIGIN
```

```
Query Match     84.4%;  Score 451.8;  DB 8;  Length 1099;
Best Local Similarity  90.3%;  Pred. No. 2.1e-92;
Matches  483;  Conservative  0;  Mismatches  52;  Indels  0;  Gaps  0;

Qy      1  CGGCCGCCTTCCTCGCCCCAGCCAGGAGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT  60
           |||||||||||||||||||||||||||||||||||||||| ||||||||| ||| ||||||
Db    200  CGGCCGCCTTCCTCGCCCCAGCCAGGAGAGCGAGATTGCAGGCATAGAGAACGACGAGGGCT  259

Qy     61  TCGGGGCCACCTGCCGCCGCCAGCCAGGTGGCCTTCTGCCCAGCCCAGGACTCGGCGAGCCGGGGTG  120
           ||||||||||||||| ||| |||||| |||||||||||| |||| ||||| ||||||||||
Db    260  TCGGGGCACCTGCCGCCAGCCATGCGGCCCCCGCCCGGGCAGCCCGCGGGCCCCACGAGTGGGGGCTG  319

Qy    121  GTTCGGAGGACATGGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTCTG  180
           ||||| ||||||||||||||| |||||||||||| ||||||||||||||||
Db    320  GTTCTGAGGACATGGGGACCACAGTCAATGAGATGTGTTTCAGGAGGCCAACCGTCCTG  379

Qy    181  CCGATGGCTACGCTCGATTGCCCAGCCGGACGAGTTGACTCAGGAGCATCC  240
           ||||| ||||||||||||||||||| ||| ||||||||||| ||||||||
Db    380  CTGATGGCTACGCACGCCAGCCATTGCCCAGGCTGACAGGCTGACCAGGAGCATCC  439

Qy    241  GCAAGTGGAGAGAGCAGAAGAAAGGCTGCAGGAGTTGGATGCTGCCTGCCGGAAGGTGA  300
           ||||||||||||||| ||||||||||||||||||||| ||||| ||||| ||||||||
Db    440  GCAAGTGGCGGGAGGAGCAGAGGAAACTGCAGGAGCTGGATGCTGCATCTAAGGTCA  499

Qy    301  CCGAACAGGAGTGGCGGGAGAAGACCTGGAACTGGAGAGTGGAACCAGCGCCAAA  360
           ||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db    500  CGGAACAGGAATGGCGGGAGAAGGCCAAGAAGGACCTGGAGGAGTGGAACCAGCGCCAGA  559
```

```
Qy      481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCGCCTGCCTCGGTGCTCAT 535
            |||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      694 ACCCCAAGAGCAGCAAGCAGTGCAAAGATGTGTCCCGCCTGCCTCGGTGCTCAT 748
```

```
RESULT 14
BC070404
LOCUS         BC070404                1009 bp    mRNA    linear  ROD 30-MAR-2005
DEFINITION    Mus musculus clathrin, light polypeptide (Lcb), mRNA (cDNA clone
              MGC:92950 IMAGE:5699653), complete cds.
ACCESSION     BC070404
VERSION       BC070404.1  GI:47682641
KEYWORDS      MGC.
SOURCE        Mus musculus (house mouse)
  ORGANISM    Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE     1  (bases 1 to 1009)
  AUTHORS     Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
              Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
              Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
              Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
              Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
              Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
              Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
              Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
              Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
              McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
              Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
              Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
              Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
              Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
              Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
              Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
              Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
              Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  CONSRTM     Mammalian Gene Collection Program Team
  TITLE       Generation and initial analysis of more than 15,000 full-length
              human and mouse cDNA sequences
  JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  PUBMED      12477932
REFERENCE     2  (bases 1 to 1009)
  AUTHORS     NIH MGC Project
  CONSRTM     Direct Submission
  TITLE       Submitted (10-MAY-2004) National Institutes of Health, Mammalian
  JOURNAL     Gene Collection (MGC), Bethesda, MD 20892-2590, USA
              NIH-MGC Project URL: http://mgc.nci.nih.gov
  REMARK      Contact: MGC help desk
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Dr. Jim Lin, University of Iowa
              cDNA Library Preparation: M. Bento Soares, University of Iowa
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
              Thomas L. Casavant.
              Web site: http://genome.uiowa.edu
  COMMENT     Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
              Bonaldo,M.F., Akabogu,I., Bair,T., Bair,J., Crouch,K., Davis,A.,
              Fishler,K., Keppel,C., Kucaba,T., Lebeck,M., Melo,A., Schaefer,K.,
              Scheetz,T., Smith,C., Snir,E., Tack,D., Trout,K., Walters,J.,
              Casavant,T., Soares,M.B.
              Clone distribution: MGC clone distribution information can be found
              through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
FEATURES             Location/Qualifiers
     source          1..1009
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="MGC:92950 IMAGE:5699653"
```

```
Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 17 Row: m Column: 15
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, GenomeScan gene prediction.

FEATURES             Location/Qualifiers
     source          1..1164
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:12930 IMAGE:4299637"
                     /tissue_type="Eye, normal, pigmented retinal epithelium"
                     /clone_lib="NIH_MGC_43"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene            1..1164
                     /gene="CLTB"
                     /note="synonym: LCB"
                     /db_xref="GeneID:1212"
                     /db_xref="MIM:118970"
     CDS             144..779
                     /gene="CLTB"
                     /codon_start=1
                     /product="clathrin, light polypeptide, isoform a"
                     /protein_id="AAH06332.1"
                     /db_xref="GI:13623459"
                     /db_xref="GeneID:1212"
                     /db_xref="MIM:118970"
                     /translation="MADDFGFFSSSESGAPEAAEBDPAAAFLAQQESEIAGIENDEGF
                     GAPAGSHAAPAQPGPTSGAGSEDMGTTVNGDVFQBANGPADGYAAIAQADRLTQEPES
                     IRKWREEQRKRLQELDAASKVTEQWREKAKKDLEEWNQRQSEQVEKNKINNRASEEA
                     FVKESKEETPGTEWEKVAQLCDFNPKSSKQCKDVSRLRSVLMSLKQTPLSR"

ORIGIN

Query Match           84.4%;  Score 451.8;  DB 8;  Length 1164;
Best Local Similarity 90.3%;  Pred. No. 2.1e-92;
Matches  483;  Conservative  0;  Mismatches  52;  Indels  0;  Gaps  0;

Qy        1 CGGCCCGCCTTCCTGGCCCAGCAGGAGGAGGAGGAGATTGCTGGCCATCGAGAATGACTCGGGTT   60
            |||||||||||||||||||||||||||||||||||| || |||||||||||||||||||||
Db      214 CGGCCGCCTTCCTGGCCCAGCAGGAGGAGGAGGAGATTGCAGGCCATAGAGAACGACGAGGGGCT  273

Qy       61 TCGGGGCACCTGCCGCCGGCCAGGTGGCTGGCCTCTGCGCACGCCCGGACTTCGGCGAGCGGGGTG  120
            |||||||| |||| |||||||||||||||||||| |||||||||||||||| ||||||||||
Db      274 TCGGGGGCACCTGCCCGGCAGGCCATGCGGCCCCGCGGCGCCCACGAGTGTGGGGGCTG         333

Qy      121 GTTCGGAGGACATGGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACGGGCCTG          180
            ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db      334 GTTCTGAGGACATGGGGACCACAGTCAATGAGATGTGTTTCAGGAGGCCAACGGGTCCTG        393

Qy      181 CCGATGGCTACGCTGCGATTGCCCAGGCGGACCAGGTTGACTCAGGAGCCTGAGAGCATCC       240
            ||||||||| ||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      394 CTGATGGCTACGCCGCCATTGCCCAGGCTGACCAGGCTGACAGCCTGAGAGCATCC         453

Qy      241 GCAAGTGGAGAGGAGGCAGAGAAAAAGCTGCAGGAGTTGATGCTGCCTCGAAGGTGA        300
            |||||||||||||||||||||||||||| |||||| |||||| |||||| ||||| |
Db      454 GCAAGTGGCGAGAGGAGGCAGAGAAACGGCTGCAAGAGCTGATGCTGCAATCTAAGGTCA       513

Qy      301 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA  360
            |||||||||||||||||||||||||||| || || ||||||||||||||||||||||||
Db      514 CGGAACAGGAATGGCGGGAGAAGGCCAAGAAGGACCTGGAGGAGTGGAACCAGCGCCAGA      573

Qy      361 GTGAACAGGTTGAGAAGAACAAGATCAACACACAGGGCATCGGAAGAGGCTTTTGTGAAAG    420
            ||||||||||||||||||||||||||||||||| |||| ||||||||||||||||||||
Db      574 GTGAACAAGTAGAGAAGAACAAGATCAACAACCACCGGGCATCCGAGAGGCTTTCGTGAAGG    633

Qy      421 AATCCAAGGAGGAGACCCCCAGCCACAGTGGCCCAGTGTGTGACTTCA                 480
            |||||||||||||||||||| |||||||||||||||||||||||||||
Db      634 AATCCAAGGAGGAGACCCCAGGCCACAGTGGGAGAAGGTGCGCCAGTCTATGTGACTTCA      693
```

```
VERSION      X04853.1  GI:513
KEYWORDS     clathrin light chain b.
SOURCE       Bos taurus (cow)
  ORGANISM   Bos taurus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
             Pecora; Bovidae; Bovinae; Bos.
REFERENCE    1 (bases 1 to 1033)
  AUTHORS    Jackson,A.P., Seow,H.F., Holmes,N., Drickamer,K. and Parham,P.
  TITLE      Clathrin light chains contain brain-specific insertion sequences
             and a region of homology with intermediate filaments
  JOURNAL    Nature 326 (6109), 154-159 (1987)
  PUBMED     3821891
REFERENCE    2 (bases 1 to 1033)
  AUTHORS    Jackson,A.P.
  TITLE      Direct Submission
  JOURNAL    Submitted (12-MAR-1987) Jackson A.P., Stanford University, Dep. of
             Cell Biology, Sherman Fairchild Building, Stanford, Ca 94305, USA
FEATURES             Location/Qualifiers
     source          1..1033
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /cell_line="BL-3"
                     /clone_lib="bovine lymphocyte cDNA"
     CDS             143..775
                     /note="unnamed protein product; lymphocyte LCb (AA 1-210)"
                     /codon_start=1
                     /protein_id="CAA28544.1"
                     /db_xref="GI:514"
                     /db_xref="GOA:P04975"
                     /db_xref="UniProt/Swiss-Prot:P04975"
                     /translation="MADDFGFFSSESGAPEAAEEDPAAAFLAQQESEIAGIENDEGF
                     GAPAGSQGGLAQPGPASGASEDMGATVNGDVFQEANGPADGYAAIAQADRLTQEPESI
                     RKWREEQKRLQELDAASKVMEQEWREKAKKDLEEWNQRSEQVEKNKINNRASEEAF
                     VKESKEETPGTEWEKVAQLCDFNPKSSKQCKDVSRLRSVLMSLKQTPLSR"
     polyA_site      1033
                     /note="polyA site"
ORIGIN
Query Match     80.3%;  Score 429.4;  DB 4;  Length 1033;
Best Local Similarity  89.0%;  Pred. No. 2.6e-87;
Matches  476;  Conservative  0;  Mismatches  56;  Indels  3;  Gaps  1;

Qy      1 CGGCCGCCTTCCTGCCCAGCAGGAGGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT   60
            |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    213 CGGCTGCCTTCTTGCCCCAGCAGGAGGAGTGAGATTGCGGGCATAGAGAATGACGAGGGTT  272

Qy     61 TCGGGGCACCTGCCGCGGGAGAAAGACCTGGCCTTCTGCGCAGCCCGACTCGCGAGCGGGGGTG  120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    273 TCGGGGCACCTGCCGCGGGAGAAAGACCTGGCCTTCTGCGCAGCCCGGCCCGGACCCGGCCCGCAGTGGG---G  329

Qy    121 GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    330 CTTCTGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCCAACGGGTCCG  389

Qy    181 CCGATGGCTACGCTGCGATTGCCCAGGCCGGACAGGTTGACTCAGGAGGCCTGAGAGCATCC  240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    390 CCGATGGCTATGCTGCAATCGCCCAGGCCGGACAGGTTGACTCAGGAGGACATCGAGAGCATCC  449

Qy    241 GCAAGTGGAGAGACCCCAGCACAGAGTGGGGAGAAAAGGCTGCAGGAGTTGGATGCGCTCCGAAGTGA  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    450 GCAAATGGAGAGAGCAGCGGAAACGCGTCGAGGAGTCGATGCGTGCCTCCAAAGTGA  509

Qy    301 CCGAACAGGAGTGGCGGGAGAAGCCAAAAAAGACCTGGAAGAGGCCAAAAAAGACCAGCGCCAAA  360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    510 TGGAGCAGGAGTGGCGGGAGAAGGCCAAGAAGGACCTGGAGGAATGGAACCAGCGCCAGA  569

Qy    361 GTGAACAGGTTGAGAGAGAACAGAATCAACACAGGGCATCGGAAGAGGCTTTTGTGAAAG  420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    570 GTGAACAAGTTGAAGAAGAGAGCCAAGAAGGCCAAGAATCAACAACAACCCGGGGCATCTGGGAGAGGCTTTTGTGGAAGG  629

Qy    421 AATCCAAGGAGGAGGAGACCCCAGCACAGAGTGGGGAGAAGGTGGCCCAGCTGTGTGACTTCA  480
```

RESULT 15
BTLCBRL
LOCUS   BTLCBRL   1033 bp   mRNA   linear   MAM 28-APR-2004
DEFINITION Bos taurus mRNA for clathrin light chain b (lymphocyte LCb).
ACCESSION   X04853 Y00265

```
gene            /tissue_type="Brain, mouse 15 dpc"
                /clone_lib="NIH BMAP_ERO"
                /lab_host="DH10B"
                /note="Vector: pYX-ASC"
                1..1009
                /gene="Cltb"
                /db_xref="GeneID:74325"
                /db_xref="MGI:1921575"
CDS             11..700
                /gene="Cltb"
                /codon_start=1
                /product="Cltb protein"
                /protein_id="AAH70404.1"
                /db_xref="GI:47682642"
                /db_xref="GeneID:74325"
                /db_xref="MGI:1921575"
                /translation="MAEDFGFFSSSESGAPEAAEEDPAAAFLAQQESEIAGIENDPGF
                GAPAASQVASAQPGLASGASGASEDMSTTVNGDVFQEANGPADGYAAIAQADRLTQEPES
                IRKWREEQKRLQELDAASKVTEQEWREKAKKDLEWNQRSEQVEKNKINRIADKA
                FYQQPDADTIGYVASEEAFVKESKEETPGTEWEKVAQLCDFNPKSSKQCKDVSRLRSV
                LMSLKQTPLSR"
ORIGIN
Query Match     82.7%;  Score 442.2;  DB 9;  Length 1009;
Best Local Similarity  87.8%;  Pred. No. 3.2e-90;
Matches  517;  Conservative  0;  Mismatches  18;  Indels  54;  Gaps  1;

Qy      1 CGGCCGCCTTCCTGCCCAGCAGGAGGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT   60
            ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     81 CGGCCGCCTTCCTGCCCAGCAGGAGGAAAGCCGAGATTGCTGGCATCGAGAGAATGACCCGGGTT  140

Qy     61 TCGGGGCACCTGCCGCGGGAGAAAGACCTGGCCTTCTGCGCAGCCCGGACTCGCGAGCGGGGGGTG  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 TCGGGGCACCTGCCGCGGGAGAAAGACCTGGCCTTCTGCGCAGCCCGCAGCCCGGACGGGGGCTG  200

Qy    121 GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG  180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 GTTCAGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG  260

Qy    181 CCGATGGCTACGCTGCGATTGCCCAGGCCGGACAGGTTGACTCAGGAGGCCTGAGAGCATCC  240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 CCGATGGCTACGCTGCGATTGCCCAGGCCGGACAGGTTGACTCAGGAGGCCGAGAGCATCC  320

Qy    241 GCAAGTGGAGAGACCCCAGCACAGAGTGGGGAGAAAAGGCTGCAGGAGTTGATGCGCTCCGAAGTGA  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GGAAGTGGAGAGAGACCCCAGCACAGAGTGGGGAGAAAAGGCCAAGAAGGCCAAGAAGGCCAAGGAGTTGATGCGCTCCGAAGTGA  380

Qy    301 CCGAACAGGAGTGGCGGGAGAAGCCAAAAAAGACCTGGAAGAGGCCAAGAACCAGCGCCAAA  360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 CCGAACAGGAGTGGCGGGAGAAGCCAAGAAGAAGACCTGGAAGAGGCCGAAGAACCAGCGCCAAA  440

Qy    361 GTGAACAGGTTGAGAAGAACAAGATCAACACAG-----------------  394
            ||||||||||||||||||||||||||||||||
Db    441 GTGAACAGGTTGAGAAGAACAAGATCAACAACAACAAGAGGATCGCTGACAAAGCGTTCTACCAGC  500

Qy    395 -----------------GGCATCGGAAGAGGCTTTTGTGAAAAGAATCCA  426
                             ||||||||||||||||||||||||||||||||
Db    501 AGCCAGATGCTGATACCATTGGCATCGGAAGAGGCTTTTGTGAAAGAATCCA  560

Qy    427 AGGAGGAGACCCCAGCACAGAGTGGGGAGAAGGTGGCCCAGCTGTGTGACTTCAAACCCTA  486
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    561 AGGAGGAGACCCCAGCACAGAGTGGGGAGAAGGTGGCCCAGCTGTGTGACTTCAATCCCA  620

Qy    487 AGAGCAGCAATGTAAAGACGTGTCCCGCGTCGGGTGCTCAT  535
            ||||||||||||||||||||||||||||||||||||||
Db    621 AGAGCAGCAAGCAGTGTAAAGATGTGTCCCGCGTCGGGTGCTCAT  669
```

Db

```
       |||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
  630  AATCCAAGGAGGAGGAGACCCCGGGCACAGAGTGGGAGAAGGTGCCCAGCTGTGTGACTTCA  689
```

Qy

```
  481  ACCCTAAGAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTGGCGTCGGTGCTCAT  535
       ||||  ||||||||||  ||||||||  |  ||||||||||||||||||||||||||
  690  ACCCCAAGAGAGCAGCAAGCAGTGCAAAGACGTGTCCCGCCTGCGTCGGTGCTCAT  744
```

Db

Search completed: January 5, 2006, 20:29:50
Job time : 2906 secs

```
                GenCore version 5.1.6
          Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    January  5, 2006, 17:19:45 ; Search time 487 Seconds
                                        (without alignments)
                                        7321.582 Million cell updates/sec

Title:          US-10-664-705-145
Perfect score:  535
Sequence:       1 cggccgccttcctggccag.........cgcctgcggctcggtgctcat 535

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       4996997 seqs, 333234630B residues

Total number of hits satisfying chosen parameters:    9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      N_Geneseq_21:*
                1: geneseqn1980s:*
                2: geneseqn1990s:*
                3: geneseqn2000s:*
                4: geneseqn2001as:*
                5: geneseqn2001bs:*
                6: geneseqn2002as:*
                7: geneseqn2002bs:*
                8: geneseqn2003as:*
                9: geneseqn2003bs:*
               10: geneseqn2003cs:*
               11: geneseqn2003ds:*
               12: geneseqn2004as:*
               13: geneseqn2004bs:*
               14: geneseqn2005s:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	535	100.0	535	13	ADR15112	Adr15112 Rat elect
2	471	88.0	982	6	ABK63500	Abk63500 Rat seque
3	471	88.0	982	10	ADB58810	Adb58810 Toxicity-
4	471	88.0	982	10	ADB53521	Adb53521 Primary r
5	471	88.0	982	10	ABT42334	Abt42334 Toxicity
6	471	88.0	982	11	ADW21848	Adw21848 Rat hepat
7	471	88.0	982	12	ADP72639	Adp72639 Renal tox
8	471	88.0	982	13	ADV41572	Adv41572 Rat cardi
9	451.8	84.4	1051	6	ABK83950	Abk83950 Human CDN
c 10	451.8	84.4	1051	12	ADJ62744	Adj62744 Human CDN
c 11	451.8	84.4	1051	14	ADX07445	Adx07445 Cyclin-de
12	387.8	72.5	1134	14	ADX07447	Adx07447 Cyclin-de
13	380.8	71.2	894	6	ABK34444	Abk34444 Human CDN
14	379.2	70.9	907	6	ABK34409	Abk34409 Human CDN
15	364.4	68.1	490	9	ACH32890	Ach32890 Human end
16	293.8	54.9	493	9	ACH39643	Ach39643 Human foe
17	268.8	50.2	485	9	ACH40125	Ach40125 Human foe
18	240.2	44.9	562	13	ADQ52771	Adq52771 Novel can
19	213.4	39.9	853	12	ADP04090	Adp04090 Human col
20	213.4	39.9	1023	13	ACN38893	Acn38893 Tumour-as
21	213.4	39.9	1164	6	ABQ54448	Abq54448 Human ova
22	213.4	39.9	1233	12	ADP04094	Adp04094 Human col
23	190.8	35.7	1262	12	ADP04095	Adp04095 Human col
24	176.6	33.0	530	14	ACL55096	Acl55096 Human col
25	176.2	32.9	958	6	ABI99595	Abi99595 Mouse iBC
26	175	32.7	528	14	ACL62179	Acl62179 Human col
27	175	32.7	1666	12	ADP04091	Adp04091 Human col
28	145.4	27.2	1958	12	ADP04093	Adp04093 Human col
29	143.4	26.8	1124	10	ADB53376	Adb53376 Primary r
30	143.4	26.8	1124	13	ADV41478	Adv41478 Rat cardi
31	139.4	26.1	527	6	ABI99585	Abi99585 Mouse iBC
32	138.6	25.9	1096	13	ADQ87410	Adq87410 Human tum
33	138.6	25.9	1105	6	ABK83951	Abk83951 Human CDN
34	138.6	25.9	1105	13	ABN95791	Abn95791 Gene #228
35	138.6	25.9	1105	13	ACN38894	Acn38894 Tumour-as
36	137.2	25.6	459	9	ACH40991	Ach40991 Human foe
37	130.8	24.4	5938	6	ABK84688	Abk84688 Human CDN
38	124.6	23.3	1542	12	ADP04092	Adp04092 Human col
39	121	22.6	735	10	ADC39087	Adc39087 Novel hum
40	115.4	21.6	473	9	ACH22936	Ach22936 Human adu
41	87.4	16.3	417	10	ADG10314	Adg10314 Mouse neu
42	78.2	14.6	1084	4	ABL29421	Abl29421 Drosophil
c 43	78.2	14.6	3201	4	ABL29420	Abl29420 Drosophil
44	72.6	13.6	484	9	ACH33747	Ach33747 Human end
45	72.6	13.6	617	13	ADU11758	Adu11758 Solid tum

ALIGNMENTS

RESULT 1
ADR15112
ID   ADR15112 standard; DNA; 535 BP.
XX
AC   ADR15112;
XX
DT   21-OCT-2004 (first entry)
XX
DE   Rat electroconvulsive seizure (EC) signature gene seqid 145.
XX
KW   neuroleptic; nootropic; antidepressant; eating disorder; tranquilizer;
KW   neuropsychiatric disorder; signature gene; electroconvulsive seizure;
KW   ECS; schizophrenia; autism; major depressive disorder; MDD;
KW   bipolar affective disorder; BAD; psychotic depression; anxiety disorder;
KW   eating disorder; attention deficit hyperactivity disorder;
KW   neuropsychiatric drug; gene; ss; rat.
XX
OS   Rattus norvegicus.
XX
PN   US2004152107-A1.
XX
PD   05-AUG-2004.
XX
PF   18-SEP-2003; 2003US-00664705.
XX
PR   18-SEP-2002; 2002US-0411718P.
PR   09-DEC-2002; 2002US-0431882P.
PR   18-JUN-2003; 2003US-0479970P.
XX
PA   (ALTA/) ALTAR C A.
PA   (LAEN/) LAENG P.
PA   (YOUN/) YOUNG T A.
PA   (CHAR/) CHARLES V D.
XX
PI   Altar CA, Laeng P, Young TA, Charles VD;
XX
DR   WPI; 2004-580183/56.
XX
PT   Identifying compound to treat neuropsychiatric disorder, by contacting
PT   cell with test compound, determining expression of signature genes by
PT   cell, comparing expression of signature genes to expression in cell not
PT   contacted with compound.

XX
PS   Claim 1; SEQ ID NO 145; 186pp; English.
XX
CC   The invention describes a method of identifying (M1) a compound to treat
CC   a neuropsychiatric disorder. The method involves contacting a cell with a
CC   test compound, determining the expression of one or more signature genes
CC   by the cell, each signature gene comprising a nucleic acid that
CC   hybridises to a nucleic acid chosen from 154 fully defined sequences (S1)
CC   of e.g., 4154, 1582, 645, 367, 405, 1080, 580, 1159, 539, 465, 2002,
CC   2469, 3165, 421, 2957, 644, 2562, 2858, 1194 and 1227 base pairs as given
CC   in the specification and their complements, and comparing determined
CC   expression of one or more signature genes to expression in cell not
CC   contacted with test compound. Also described are: selecting (M2) one or
CC   more signature genes that are indicative of an effective therapy for
CC   treating a neuropsychiatric disorder; and a kit (I1) for detecting an
CC   electroconvulsive seizure (ECS) gene signature, comprising several
CC   oligonucleotides, each of which is capable of specifically hybridising to
CC   a different ECS signature gene. Therapeutic methods, which use ECS gene
CC   signatures and ECS signature genes to treat, ameliorate or prevent
CC   neuropsychiatric disorders, is also disclosed. (M1) is useful for
CC   identifying a compound to treat a neuropsychiatric disorder chosen from
CC   schizophrenia, autism, major depressive disorder (MDD), bipolar affective
CC   disorder (BAD) and psychotic depression. (M1) is useful for identifying
CC   compounds for treating neuropsychiatric disorders such as anxiety
CC   disorders, eating disorders and attention deficit hyperactivity disorder.
CC   The ECS signature genes of (M1) are useful in diagnostic and prognostic
CC   methods of identifying people who have or who are susceptible to
CC   neuropsychiatric disorders such as MDD and BAD. (M1) enables multi-
CC   parameter high throughput screening of compounds for treating
CC   neuropsychiatric disorders; screening of large number of candidate
CC   compounds for treating neuropsychiatric disorder and enables cost-
CC   effective and rapid screening of compounds useful as neuropsychiatric
CC   drugs. This sequence represents an electroconvulsive seizure (ECS)
CC   signature gene.
XX
SQ   Sequence 535 BP; 138 A; 127 C; 190 G; 80 T; 0 U; 0 Other;

Query Match   100.0%;   Score 535;   DB 13;   Length 535;
Best Local Similarity   100.0%;   Pred. No. 1e-132;
Matches 535;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
Qy    1 CGGCCGGCCTTCCTGGCCCAGCAGGAGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 CGGCCGGCCTTCCTGGCCCAGCAGGAGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT   60

Qy   61 TCGGGGCACCTGCCGCGCCCAGGTGGCCTCTGCGCAGCCCGGACTCGCGAGCGGGGGGTG  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TCGGGGCACCTGCCGCGCCCAGGTGGCCTCTGCGCAGCCCGGACTCGCGAGCGGGGGGTG  120

Qy  121 GTTCGGAGGACATGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACGGGCCTG  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GTTCGGAGGACATGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACGGGCCTG  180

Qy  181 CCGATGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCCTGAGAGCATCC  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CCGATGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCCTGAGAGCATCC  240

Qy  241 GCAAGTGGAGAGGAGGACAGAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GCAAGTGGAGAGGAGGACAGAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA  300

Qy  301 CCGAACAGGAGTGCGGGGAGAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 CCGAACAGGAGTGCGGGGAGAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA  360

Qy  361 GTGAACAGGTTGAAGAAGAACAAGATCAACAACACAGGGCATCGGAAGAGGCTTTTGTGAAAG  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 GTGAACAGGTTGAAGAAGAACAAGATCAACAACACAGGGCATCGGAAGAGGCTTTTGTGAAAG  420

Qy  421 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCA  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCA  480
```

Qy   481 ACCCTAAGAGCAGCAGCAATGCAAGCAATGTAAAGACGTGTCCGCCTGCGGTGCTCAT 535
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ACCCTAAGAGCAGCAGCAATGCAAGCAATGTAAAGACGTGTCCGCCTGCGGTGCTCAT 535

RESULT 2
ABK63500
ID   ABK63500 standard; cDNA; 982 BP.
XX
AC   ABK63500;
XX
DT   18-JUN-2002   (first entry)
XX
DE   Rat sequence differentially expressed in response to a hepatotoxin #1407.
XX
KW   Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
KW   differential expression; centrilobular necrosis; steatosis.
XX
OS   Rattus norvegicus.
XX
PN   WO200210453-A2.
XX
PD   07-FEB-2002.
XX
PF   30-JUL-2001; 2001WO-US023872.
XX
PR   31-JUL-2000; 2000US-0222040P.
PR   02-NOV-2000; 2000US-0244880P.
PR   11-MAY-2001; 2001US-0290029P.
PR   15-MAY-2001; 2001US-0290645P.
PR   22-MAY-2001; 2001US-0292336P.
PR   06-JUN-2001; 2001US-0295798P.
PR   13-JUN-2001; 2001US-0297457P.
PR   19-JUN-2001; 2001US-0298884P.
PR   09-JUL-2001; 2001US-0303459P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;
XX
DR   WPI; 2002-241625/29.
XX
PT   Predicting toxic effects of compounds or the progression of these toxic
PT   effects by determining the changes in gene expression in tissues or cells
PT   exposed to the toxin and comparing these to gene expression in unexposed
PT   tissues or cells.
XX
PS   Claim 1; SEQ ID NO 1407; 239pp; English.
XX
CC   The invention relates to methods for predicting toxic effects of
CC   compounds or the progression of these toxic effects by determining the
CC   global changes in gene expression in tissues or cells exposed to the
CC   toxin and comparing these to gene expression in unexposed tissues or
CC   cells. Also included are methods of predicting at least one toxic effect
CC   of a compound or progression of a toxic effect, preferably the
CC   hepatotoxicity of a compound, comprising detecting the level of
CC   expression in a tissue or cell sample exposed to the compound of two or
CC   more genes listed in the specification, where differential expression of
CC   the genes is indicative of at least one toxic effect or progression. The
CC   method can also be used to identify an agent which modulates the toxic
CC   response and predict cellular pathways that a compound modulates in a
CC   cell. The methods utilise a set of at least two probes (on a solid
CC   support in kit form), where each of the probes comprises a sequence that
CC   specifically hybridises to a gene listed in the specification, a computer
CC   system comprising a database containing information identifying the
CC   expression level in a tissue or cell sample exposed to a hepatotoxin of a
CC   set of genes comprising at least two genes listed in the specification,
CC   and a user interface to view the information used to present information
CC   identifying the expression level in a tissue or cell of at least one gene
CC   listed in the specification. The method is useful for elucidating global
CC   changes in gene expression and for identifying toxicity markers in
CC   tissues or cell exposed to a known toxin. The genes may be used as
CC   toxicity markers in drug screening and toxicity assays. The genes and

CC gene expression information may be used as diagnostic markers for the
CC prediction or identification of the physiological state of tissue or cell
CC sample that has been exposed to a compound or agent. Hepatotoxicity is
CC characterized by centrilobular necrosis and steatosis. The present
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene
CC which is differentially expressed in response to a hepatotoxic agent
XX
SQ Sequence 982 BP; 212 A; 266 C; 315 G; 189 T; 0 U; 0 Other;

Query Match   88.0%; Score 471; DB 6; Length 982;
Best Local Similarity   90.8%; Pred. No. 1.5e-115;
Matches 535; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

```
Qy    1 CGGCCGCCTTCCTGGCCCAGCAGGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   99 CGGCCGCCTTCCTGGCCCAGCAGGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT 158

Qy   61 TCGGGGCACCTGCCGCGCAGGTGGCCTCTCGCGCAGCCCGGACTCGCGAGCGGGGGGTG 120
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  159 TCGGGGCACCTGCCGCGCAGGTGGCCTCTCGCGCAGCCCGGACTCGCGAGCGGGGGGTG 218

Qy  121 GTTCGGGAGGACATGGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACGGCCTG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  219 GTTCGGGAGGACATGGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACGGCCTG 278

Qy  181 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGGCCTGAGAGCATCC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  279 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGGCCTGAGAGCATCC 338

Qy  241 GCAAGTGGAGGAGGAGGAGCAGAAGAAGAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  339 GCAAGTGGAGGAGGAGGAGCAGAAGAAGAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA 398

Qy  301 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGAACCAGCGCCCAAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  399 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGAACCAGCGCCCAAA 458

Qy  361 GTGAACAGGTTGAGAAGAACAAGAATCAACACACAG------------------------ 394
          ||||||||||||||||||||||||||||||||||||
Db  459 GTGAACAGGTTGAGAAGAACAAGAATCAACACACAGCTGACAAAGCGTTCTACCAGC 518

Qy  395 ------------GGCATCGGAAGAGAGGCTTTTGTGAAAGAATCCA 426
                      ||||||||||||||||||||||||||||||||||
Db  519 AGCCAGATGCTGATACCATTGGCTATGTGCATCGGAAGAGAGGCTTTTGTGAAAGAATCCA 578

Qy  427 AGGAGGAGACCCCAGGCACAGAGGTGGGAGAAGGTTGGCCCAGCTGTGTGACTTCAACCCTA 486
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  579 AGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTA 638

Qy  487 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTTGCGCTTCGGTGCTCAT 535
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  639 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTTGCGCTTCGGTGCTCAT 687
```

RESULT 3
ADB58810
ID ADB58810 standard; DNA; 982 BP.
XX
AC ADB58810;
XX
DT 04-DEC-2003 (first entry)
XX
DE Toxicity-related gene, SEQ ID 3836.
XX
KW Toxic; toxin; gene expression profile; hepatotoxicity; liver;
KW drug screening; toxicity assay; ds.
XX
OS Unidentified.
XX
PN WO2003064624-A2.
XX
PD 07-AUG-2003.
XX

PF 31-JAN-2003; 2003WO-US003194.
XX
PR 31-JAN-2002; 2002US-00060087.
PR 15-MAR-2002; 2002US-0364045P.
PR 15-MAR-2002; 2002US-0364055P.
PR 30-DEC-2002; 2002US-0436643P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
XX
DR WPI; 2003-689530/65.
XX
PT Predicting a toxic effect of a compound, useful in identifying toxicity
PT markers in liver tissues or cells for drug screening and toxicity assays,
PT comprises preparing gene expression profile of tissue or cells exposed to
PT the compound.
XX
PS Claim 1; SEQ ID NO 3836; 1156pp; English.
XX
CC The present invention relates to a method for predicting a toxic effect
CC of a compound. The method comprises preparing a gene expression profile
CC of a tissue or cell sample exposed to the compound, and comparing the
CC gene expression profile to a database comprising SEQ ID 1-4925, where
CC differential expression of the gene indicates at least one toxic effect.
CC The method is useful for predicting at least one toxic effect of a
CC compound, predicting hepatotoxicity or the progression of a toxic effect
CC of a compound, identifying an agent that modulates the onset or
CC progression of a toxic response, predicting the cellular pathways that a
CC compound modulates in a cell, and identifying an agent that modulates at
CC least one activity of a protein. The method and compositions of the
CC present invention using a database of genes having liver toxin-induced
CC differential expression, are useful in identifying toxicity markers in
CC liver tissues or cells for drug screening and toxicity assays. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 982 BP; 212 A; 266 C; 315 G; 189 T; 0 U; 0 Other;

Query Match   88.0%; Score 471; DB 10; Length 982;
Best Local Similarity   90.8%; Pred. No. 1.5e-115;
Matches 535; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

```
Qy    1 CGGCCGCCTTCCTGGCCCAGCAGGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   99 CGGCCGCCTTCCTGGCCCAGCAGGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT 158

Qy   61 TCGGGGCACCTGCCGCGCAGGTGGCCTCTCGCGCAGCCCGGACTCGCGAGCGGGGGGTG 120
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  159 TCGGGGCACCTGCCGCGCAGGTGGCCTCTCGCGCAGCCCGGACTCGCGAGCGGGGGGTG 218

Qy  121 GTTCGGGAGGACATGGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACGGCCTG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  219 GTTCGGGAGGACATGGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACGGCCTG 278

Qy  181 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGGCCTGAGAGCATCC 240
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  279 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGGCCTGAGAGCATCC 338

Qy  241 GCAAGTGGAGGAGGAGGAGCAGAAGAAGAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  339 GCAAGTGGAGGAGGAGGAGCAGAAGAAGAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA 398

Qy  301 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGAACCAGCGCCCAAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  399 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGAACCAGCGCCCAAA 458

Qy  361 GTGAACAGGTTGAGAAGAACAAGAATCAACACACAG------------------------ 394
          ||||||||||||||||||||||||||||||||||||
Db  459 GTGAACAGGTTGAGAAGAACAAGAATCAACACACAGCTGACAAAGCGTTCTACCAGC 518

Qy  395 ------------GGCATCGGAAGAGAGGCTTTTGTGAAAGAATCCA 426
```

```
CC      effect, particularly hepatotoxicity, of a test or unknown compound. The
CC      genes listed in the specification are useful as diagnostic or toxicity
CC      markers for the prediction or identification of the physiological state
CC      of tissue or cell sample that has been exposed to a compound, or to
CC      identify or predict the toxic effects of a compound or an agent. These
CC      may also be used as markers for monitoring toxicity progression or for
CC      drug screening. The present sequence represents a primary rat hepatocyte
CC      toxicity modelling related gene sequence from the present invention.
XX
SQ      Sequence 982 BP; 212 A; 266 C; 315 G; 189 T; 0 U; 0 Other;
```

Query Match   88.0%;   Score 471;   DB 10;   Length 982;
Best Local Similarity   90.8%;   Pred. No. 1.5e-115;
Matches 535;   Conservative   0;   Mismatches   0;   Indels   54;   Gaps   1;

```
Qy      1  CGGCCGCCTTCCTGCCGCCAGCCAGGAGAGGAGGAGGAGGAGGAGATTGCTGGCATCGAGAATGACTCGGGTT   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     99  CGGCCGCCTTCCTGCCGCCAGCCAGGAGAGGAGGAGGAGGAGGAGATTGCTGGCATCGAGAATGACTCGGGTT  158

Qy     61  TCGGGGCACCTGCCGCCGCCAGCCAGGTGGCCTCTGCCGCAGCCCGACTCGCGAGCGGGGGTG  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    159  TCGGGGCACCTGCCGCCGCCAGCCAGGTGGCCTCTGCCGCAGCCCGACTCGCGAGCGGGGGTG  218

Qy    121  GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219  GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG  278

Qy    181  CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGGCTGAGAGCATCC  240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    279  CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGGCTGAGAGCATCC  338

Qy    241  GCAAGTGGAGAGGAGAGGAGCAGAGAAAAGGCTGCAGGAGTTGATGCTGCCTCGAAGGTGA  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    339  GCAAGTGGAGAGGAGAGGAGCAGAGAAAAGGCTGCAGGAGTTGATGCTGCCTCGAAGGTGA  398

Qy    301  CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    399  CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA  458

Qy    361  GTGAACAGGTTGAGAAGAACAAGATCAACAACAG---------------------------  394
           |||||||||||||||||||||||||||||||||
Db    459  GTGAACAGGTTGAGAAGAACAAGATCAACAACAGGAGATCGCTGACAAAGCGTTCTACCAGC  518

Qy    395  ---------GGCATCGGAAGAGAGCGTTTTGTGAAAGAATCCA  426
                    |||||||||||||||||||||||||||||||||
Db    519  AGCCAGATGCTGATACCATTGGCTATGTGGCATCGGAAGAGAGCGTTTTGTGAAAGAATCCA  578

Qy    427  AGGAGGAGACCCCAGGCACAGAGTGGGAGAAGTGGCCCAGCTGTGTGACTTCAACCCTA  486
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    579  AGGAGGAGACCCCAGGCACAGAGTGGGAGAAGTGGCCCAGCTGTGTGACTTCAACCCTA  638

Qy    487  AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTCGGCTCGGTGCTCAT  535
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db    639  AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTGCGCTCGGTGCTCAT  687
```

```
Db    519  AGCCAGATGCTGATACCATTGGCTATGTGGCATCGGAAGAGAGCGTTTTGTGAAAGAATCCA  578
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    427  AGGAGGAGACCCCAGGCACAGAGTGGGAGAAGTGGCCCAGCTGTGTGACTTCAACCCTA  486
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    579  AGGAGGAGACCCCAGGCACAGAGTGGGAGAAGTGGCCCAGCTGTGTGACTTCAACCCTA  638

Qy    487  AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTCGGCTCGGTGCTCAT  535
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db    639  AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTGCGCTCGGTGCTCAT  687
```

```
RESULT 4
ADB53521
ID      ADB53521 standard; DNA; 982 BP.
XX
AC      ADB53521;
XX
DT      04-DEC-2003 (first entry)
XX
DE      Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4063.
XX
KW      toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;
KW      toxicity marker; toxicity progression; drug screening;
KW      primary rat hepatocyte toxicity modelling; gene; ds.
XX
OS      Rattus norvegicus.
XX
PN      WO200306 5993-A2.
XX
PD      14-AUG-2003.
XX
PF      04-FEB-2003; 2003WO-US003482.
XX
PR      04-FEB-2002; 2002US-0353171P.
PR      13-MAR-2002; 2002US-0363534P.
PR      08-APR-2002; 2002US-0370248P.
PR      10-APR-2002; 2002US-0371134P.
PR      10-APR-2002; 2002US-0371135P.
PR      10-APR-2002; 2002US-0371150P.
PR      11-APR-2002; 2002US-0371413P.
PR      19-APR-2002; 2002US-0373601P.
PR      19-APR-2002; 2002US-0373602P.
PR      22-APR-2002; 2002US-0374139P.
PR      08-MAY-2002; 2002US-0378370P.
PR      09-MAY-2002; 2002US-0378652P.
PR      09-MAY-2002; 2002US-0378653P.
PR      09-MAY-2002; 2002US-0378665P.
PR      09-JUL-2002; 2002US-0394230P.
PR      09-JUL-2002; 2002US-0394253P.
PR      04-SEP-2002; 2002US-0407688P.
PR      28-JAN-2003; 2003US-0442900P.
XX
PA      (GENE-) GENE LOGIC INC.
XX
PI      Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;
PI      Elashoff M;
XX
DR      WPI; 2003-731472/69.
XX
PT      Determining if a compound induces a toxic effect on a tissue or cell, for
PT      identifying hepatotoxic compounds, comprises comparing a gene expression
PT      profile of a tissue or cell sample to a database of Tox mean and non-Tox
PT      mean values.
XX
PS      Claim 44; SEQ ID NO 4063; 874pp; English.
XX
CC      The present invention describes a method for determining whether a
CC      compound induces a toxic effect on a tissue or cell. The method comprises
CC      preparing a gene expression profile of a tissue or cell sample exposed to
CC      the compound, and comparing the gene expression profile to a database
CC      comprising data or information on the Tox mean and non-Tox mean value.
CC      The method is useful for predicting or identifying at least one toxic
```

```
Db    519  AGCCAGATGCTGATACCATTGGCTATGTGGCATCGGAAGAGAGCGTTTTGTGAAAGAATCCA  578
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    427  AGGAGGAGACCCCAGGCACAGAGTGGGAGAAGTGGCCCAGCTGTGTGACTTCAACCCTA  486
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    579  AGGAGGAGACCCCAGGCACAGAGTGGGAGAAGTGGCCCAGCTGTGTGACTTCAACCCTA  638

Qy    487  AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTCGGCTCGGTGCTCAT  535
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db    639  AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTGCGCTCGGTGCTCAT  687
```

RESULT 5
ABT42334
ID   ABT42334 standard; DNA; 982 BP.
XX
AC   ABT42334;
XX
DT   26-JUN-2003 (first entry)
XX
DE   Toxicity modelling related rat gene SEQ ID No 2036.
XX
KW   Toxic effect; gene expression profile; renal toxicity; toxicity marker;
KW   database; drug screening; toxicity assay; rat; ds.
XX
OS   Rattus norvegicus.
XX
PN   WO200295000-A2.
XX

```
PD   28-NOV-2002.
XX
PF   22-MAY-2002; 2002WO-US016173.
XX
PR   22-MAY-2001; 2001US-0292335P.
PR   13-JUN-2001; 2001US-0297523P.
PR   19-JUN-2001; 2001US-0298925P.
PR   10-JUL-2001; 2001US-0303807P.
PR   10-JUL-2001; 2001US-0303808P.
PR   10-JUL-2001; 2001US-0303810P.
PR   28-AUG-2001; 2001US-0315047P.
PR   27-SEP-2001; 2001US-0324928P.
PR   22-OCT-2001; 2001US-0330462P.
PR   01-NOV-2001; 2001US-0330867P.
PR   21-NOV-2001; 2001US-0331805P.
PR   06-DEC-2001; 2001US-0336144P.
PR   19-DEC-2001; 2001US-0340873P.
PR   21-FEB-2002; 2002US-0357842P.
PR   21-FEB-2002; 2002US-0357843P.
PR   21-FEB-2002; 2002US-0357844P.
PR   15-MAR-2002; 2002US-0364134P.
PR   08-APR-2002; 2002US-0370144P.
PR   08-APR-2002; 2002US-0370206P.
PR   08-APR-2002; 2002US-0370247P.
PR   17-APR-2002; 2002US-0372794P.
PR   21-APR-2002; 2002US-0371679P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
XX
DR   WPI; 2003-148464/14.
XX
PT   Predicting at least one toxic effect of a compound, useful for toxicity
PT   modeling, comprises preparing a gene expression profile of a tissue or
PT   cell sample exposed to the compound, and comparing the gene expression
PT   profile to a database.
XX
PS   Example 4; Page; 446pp; English.
XX
CC   The invention relates to a novel method of predicting at least one toxic
CC   effect of a compound. The method comprises a gene expression profile of a
CC   tissue or cell sample exposed to the compound, and comparing the gene
CC   expression profile to a database comprising at least part of the data or
CC   information given in the specification. The methods are useful for
CC   predicting at least one toxic effect of a compound, predicting the
CC   progression of a toxic effect of a compound, predicting the renal
CC   toxicity of a compound, or identifying toxicity markers in tissues or
CC   cells exposed to known renal toxin. The genes are useful as toxicity
CC   markers in drug screening and toxicity assays, in monitoring disease or
CC   physiological states, or disease progression. This polynucleotide
CC   represents a rat DNA sequence relating to the toxic effect database
CC   described in the specification. NOTE: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from the World Intellectual Property
CC   Organization
XX
SQ   Sequence 982 BP; 212 A; 266 C; 315 G; 189 T; 0 U; 0 Other;

Query Match             88.0%;  Score 471;  DB 10;  Length 982;
Best Local Similarity   90.8%;  Pred. No. 1.5e-115;
Matches 535;   Conservative    0;  Mismatches    0;  Indels   54;  Gaps    1;

Qy        1 CGGGCGCCTTCCTGCCCAGCCAGGAGGAGGAGAATGCTGCGCATCGAGAATGACTCGGGTT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       99 CGGCGGCGCCTTCCTGCCCAGCCAGGAGGAGGAGAATGCTGCGCATCGAGAATGACTCGGGTT 158

Qy       61 TCGGGGCACCTGCCCGGCCAGCCAGGTGGCCTTCGCCAGCCCGGACTCGGCGGGGGGGGTG 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      159 TCGGGGCACCTGCCCGGCCAGCCAGGTGGCCTTCGCCAGCCCGGACTCGGCGAGCGGGGTG 218

Qy      121 GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG
            ||||||||||||||||||||
```

```
Db      219 GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG 278

Qy      181 CCGATGGCTACGCTGCGATTGCCCAGGCGACAGTTGACTCAGGAGCCTGAGAGCATCC 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      279 CCGATGGCTACGCTGCGATTGCCCAGGCGACAGTTGACTCAGGAGCCTGAGAGCATCC 338

Qy      241 GCAAGTGGAGAGAGAGGAGCAGAGAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      339 GCAAGTGGAGAGAGAGGAGCAGAGAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA 398

Qy      301 CCGAACAGAGAGTGCGCGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      399 CCGAACAGAGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 458

Qy      361 GTGAACAGGTTGAGGAAGAACAAGATCAACAACAG--------------------- 394
            |||||||||||||||||||||||||||||||||||
Db      459 GTGAACAGGTTGAGGAAGAACAACAGATCAACAACAGCCTGACAAAGGCGTTCTACCAGC 518

Qy      395 ---------------------GCATCGGAAGGAGGCTTTGTGAAAGAATCCA 426
                                 ||||||||||||||||||||||||||||||
Db      519 AGCCAGATGCTGATACCATTGGCTATGTGGCATCGGAAGAGGCTTTGTGAAAGAATCCA 578

Qy      427 AGGAGGACACCCAGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTA 486
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      579 AGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTA 638

Qy      487 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTGCGCTCGGTGCTCAT 535
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db      639 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTGCGCTCGGTGCTCAT 687

RESULT 6
ADW21848
ID   ADW21848 standard; cDNA; 982 BP.
XX
AC   ADW21848;
XX
DT   10-MAR-2005  (first entry)
XX
DE   Rat hepatotoxicity marker gene, SEQ:227.
XX
KW   Toxicology screening; drug screening; gene expression;
KW   expression profile; hepatotoxicity; drug-induced; hepatitis;
KW   liver disease; gastrointestinal disease; gene; ss.
XX
OS   Rattus norvegicus.
XX
PN   EP1344834-A2.
XX
PD   17-SEP-2003.
XX
PF   04-MAR-2003; 2003EP-00004810.
XX
PR   14-MAR-2002; 2002EP-00005336.
PR   17-JUL-2002; 2002EP-00015657.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Boess F, Suter-Dick L, Wolf D;
XX
DR   WPI; 2003-723475/69.
DR   GENBANK; M15883.
XX
PT   Predicting toxicity of compounds, useful in development of safe drugs, by
PT   measuring the differential expression of specific genes in cells exposed
PT   to test compounds.
XX
PS   Claim 2; SEQ ID NO 227; 895pp; English.
XX
CC   The invention relates to methods of predicting at least one toxic effect
CC   (or toxicity progression or the mechanism of toxicity) of a compound. The
CC   methods involve detecting the level of expression of at least one of a
CC   set of 680 genes ADW21622-ADW22301 or at least one of a set of 17 genes
```

```
CC   (including ADW22362, ADW22414 and ADW22481-ADW22483) in a tissue or cell
CC   exposed to the compound, and determining whether the gene is
CC   differentially expressed compared with a control tissue or cell.
CC   Differential expression of the gene in the presence of the compound is
CC   indicative of a toxic effect, of toxicity progression or of a specific
CC   mechanism of toxicity. The toxic effect is especially hepatotoxicity,
CC   particularly hepatitis, liver necrosis, protein adduct formation or fatty
CC   liver. The invention also relates to sets of primers and probes specific
CC   for at least two genes selected from ADW21622-ADW22301; solid supports
CC   (e.g., DNA chips) and kits containing the probes; and a database
CC   containing DNA sequence information and expression information for at
CC   least two of the 680 genes from hepatotoxin-exposed tissues. The
CC   invention is based on the determination of global changes in gene
CC   expression in tissues or cells exposed to known toxins, particularly
CC   hepatotoxins, and the identification of individual genes (toxicity
CC   markers) that are differentially expressed on toxin exposure. The changes
CC   in gene expression can be characteristic of different mechanisms of
CC   hepatotoxicity mediated by various classes of compounds. Such compounds
CC   include: direct acting compounds which cause damage to macromolecules,
CC   especially proteins and lipids by directly interacting with them;
CC   steatotic compounds which cause an accumulation of fat in the liver; and
CC   cholestatic compounds which impair bile flow or bile acid transport,
CC   resulting in jaundice. The methods of the invention are useful in
CC   toxicology screening for predicting the toxic effects (especially
CC   hepatotoxic effects) of compounds for the development of safer drugs.
CC   Sequences ADW21622-ADW22301 represent specifically claimed hepatotoxicity
CC   marker genes of rat origin whose expression is altered on exposure to
CC   hepatotoxins.
XX
SQ   Sequence 982 BP; 212 A; 266 C; 315 G; 189 T; 0 U; 0 Other;

  Query Match             88.0%;  Score 471;  DB 11;  Length 982;
  Best Local Similarity   90.8%;  Pred. No. 1.5e-115;
  Matches  535;  Conservative    0;  Mismatches    0;  Indels   54;  Gaps    1;

Qy          1 CGGCCGCCTTCCTGGCCCAGCAGGAGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 CGGCCGCCTTCCTGGCCCAGCAGGAGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT 158

Qy         61 TCGGGGCACCTGCCGCCAGCCAGGTGGCCTCTGCGCAGCCCGGACTCGCGAGCGGGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 TCGGGGCACCTGCCGCCAGCCAGGTGGCCTCTGCGCAGCCCGGACTCGCGAGCGGGGGTG 218

Qy        121 GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG 278

Qy        181 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCCTGAGAGCATCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCCTGAGAGCATCC 338

Qy        241 GCAAGTGGAGAGAGGAGCAGAAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 GCAAGTGGAGAGAGGAGCAGAAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA 398

Qy        301 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 458

Qy        361 GTGAACAGGTTGAGAAGAACAAGATCAACAACAG-------------------------- 394
              ||||||||||||||||||||||||||||||||||
Db        459 GTGAACAGGTTGAGAAGAACAAGATCAACAACAGGATCGCTGACAAAGCGTTCTACCAGC 518

Qy        395 ----------------------------GGCATCGGAAGAGGCTTTTGTGAAAGAATCCA 426
                                          ||||||||||||||||||||||||||||||||
Db        519 AGCCAGATGCTGATACCATTGGCTATGTGGCATCGGAAGAGGCTTTTGTGAAAGAATCCA 578

Qy        427 AGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTA 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        579 AGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTA 638

Qy        487 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTGCGCTCGGTGCTCAT 535
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        639 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTGCGCTCGGTGCTCAT 687

RESULT 7
ADP72639
ID   ADP72639 standard; DNA; 982 BP.
XX
AC   ADP72639;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Renal toxin progression gene marker #1228.
XX
KW   ds; toxic effect; gene expression profile; kidney tissue;
KW   differential gene expression; toxicity progression; toxicity marker;
KW   drug screening; toxicity assay; kidney pathology; nephritis;
KW   kidney necrosis; glomerular injury; tubular injury;
KW   focal segmental glomerulosclerosis.
XX
OS   Rattus norvegicus.
XX
PN   WO2004048598-A2.
XX
PD   10-JUN-2004.
XX
PF   24-NOV-2003; 2003WO-US037556.
XX
PR   22-NOV-2002; 2002US-00301856.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Mendrick DL,  Porter MW,  Johnson KR,  Castle A,  Higgs B;
PI   Elashoff M;
XX
DR   WPI; 2004-460771/43.
XX
PT   Predicting (the progression of) a toxic effect of a compound, for
PT   monitoring the progression of renal disease states, comprises preparing a
PT   gene expression profile of a kidney tissue or cell sample exposed to the
PT   compound.
XX
PS   Claim 11; SEQ ID NO 1228; 266pp; English.
XX
CC   The invention relates to a method of predicting (the progression of) a
CC   toxic effect of a compound by preparing a gene expression profile of a
CC   kidney tissue or cell sample exposed to the compound and comparing the
CC   gene expression profile to a database, or detecting the level of gene(s)
CC   expression in a tissue or cell sample exposed to the compound, where
CC   differential gene expression compared to a control indicates a toxic
CC   effect (toxicity progression). The method is useful for predicting (the
CC   progression of) at least one toxic effect of a compound. The genes are
CC   useful as toxicity markers in drug screening and toxicity assays. The
CC   methods are useful for predicting the likelihood that a compound or test
CC   agent will induce various specific kidney pathologies, such as nephritis,
CC   kidney necrosis, glomerular and tubular injury, or focal segmental
CC   glomerulosclerosis. The methods are useful for determining the similarity
CC   of a toxic response to one or more individual compounds and for
CC   predicting or elucidating the potential cellular pathways influenced,
CC   induced or modulated by the compound or test agent. The kit is useful for
CC   predicting or modelling the toxic response of a test compound, for
CC   monitoring the progression of renal disease states, for identifying genes
CC   that show promise as new drug targets and for screening known and newly
CC   designed drugs. This sequence corresponds to a gene marker used in the
CC   method of the invention. (Note: The sequence data for this patent did not
CC   form part of the printed specification, but was obtained in electronic
CC   format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences).
XX
SQ   Sequence 982 BP; 212 A; 266 C; 315 G; 189 T; 0 U; 0 Other;

  Query Match             88.0%;  Score 471;  DB 12;  Length 982;
  Best Local Similarity   90.8%;  Pred. No. 1.5e-115;
  Matches  535;  Conservative    0;  Mismatches    0;  Indels   54;  Gaps    1;
```

XX
PT Predicting a toxic effect of a compound by obtaining a gene expression
PT profile of a tissue or cell sample exposed to the compound and comparing
PT the gene expression profile to a database comprising all of the data
PT given in the specification.
XX
PS Example 1; SEQ ID NO 2438; 520pp; English.
XX
CC The invention relates to a method of predicting at least one toxic effect
CC of a compound by obtaining a gene expression profile of a tissue or cell
CC sample exposed to the compound and comparing the gene expression profile
CC to a database of toxicity prediction genes which are differentially
CC expressed on exposure to known toxins. The tissue or cell sample is
CC preferably derived from heart tissue, and the predicted toxic effect is
CC especially a cardiotoxic effect such as myocarditis, arrhythmias,
CC tachycardia, myocardial ischemia, angina, hypertension, hypotension,
CC dyspnea, cardiogenic shock or other specific heart pathologies. The
CC invention is based on the elucidation of global changes in gene
CC expression in tissues or cells exposed to known toxins, particularly
CC cardiotoxins, and the identification of individual genes (toxicity
CC markers) that are differentially expressed on toxin exposure. The
CC invention also relates to methods for predicting the progression of a
CC compound's toxic effects, for predicting the cardiotoxicity of a
CC compound, for identifying an agent that modulates the onset or
CC progression of a toxic response, and for predicting which cellular
CC pathways a particular compound will modulate. The invention further
CC relates to sets of at least two hybridization probes specific for
CC toxicity prediction genes, and solid supports and kits comprising them.
CC The method of the invention is useful in toxicology screening for
CC predicting the toxic effects (especially cardiotoxic effects) of
CC compounds such as pharmaceutical agents or environmental pollutants.
CC Sequences ADV39135-ADV41830 represent cardiotoxicity marker
CC polynucleotides of rat origin whose expression is altered on exposure to
CC at least one cardiotoxin. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 982 BP; 212 A; 266 C; 315 G; 189 T; 0 U; 0 Other;

Query Match   88.0%;   Score 471;   DB 13;   Length 982;
Best Local Similarity   90.8%;   Pred. No. 1.5e-115;
Matches 535;   Conservative   0;   Mismatches   0;   Indels   54;   Gaps   1;

```
Qy    1 CGGCCGCCTTCCTGGCCCAGCAGGAGGAGGAGATTGCTGGCATCGAGAATGACTCGGGTT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   99 CGGCCGCCTTCCTGGCCCAGCAGGAGGAGGAGATTGCTGGCATCGAGAATGACTCGGGTT 158

Qy   61 TCGGGGCACCTGCGCGCCAGGTGGCCTCTGCGCAGCCCGGACTCGCGAGCCGAGCCGGGTG 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  159 TCGGGGCACCTGCGCGCCAGGTGGCCTCTGCGCAGCCCGGACTCGCGAGCCGAGCCGGGTG 218

Qy  121 GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTCAGGAGGCTAACGGGCCTG 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  219 GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTCAGGAGGCTAACGGGCCTG 278

Qy  181 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGTGGTTGACTCAGGAGGCCTGAGAGCATCC 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  279 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGTGGTTGACTCAGGAGGCCTGAGAGCATCC 338

Qy  241 GCAAGTGGAGAGAGGAGCAGGAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  339 GCAAGTGGAGAGAGGAGCAGGAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA 398

Qy  301 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  399 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 458

Qy  361 GTGAACAGGTTGAGAAGAACAAGATCAACAACAG------------------------ 394
        |||||||||||||||||||||||||||||||
Db  459 GTGAACAGGTTGAGAAGAACAAGATCAACAACAGGATCCGCTGACAAAGCGTTCTACCAGC 518

Qy  395 ------------GGCATCGGAAGAGGCTTTTGTGAAAGAATCCA 426
                    ||||||||||||||||||||||||||||||||
Db  519 AGCCAGATGCTGATACCATTGGCTATGTGCATCGGAAGAGGCTTTTGTGAAAGAATCCA 578

Qy  427 AGGAGGAGACCCCAGGCACAGAGTGGGAAGTGGCCCAGTTGTGTGACTTCAACCCTA 486
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  579 AGGAGGAGACCCCAGGCACAGAGTGGGAAGTGGCCCAGTTGTGTGACTTCAACCCTA 638

Qy  487 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCGCTCGGTGCTCAT 535
        |||||||||||||||||||||||||||||||||||||||||||||
Db  639 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCGCTCGGTGCTCAT 687
```

RESULT 8
ADV41572
ID   ADV41572 standard; cDNA; 982 BP.
XX
AC   ADV41572;
XX
DT   10-FEB-2005 (first entry)
XX
DE   Rat cardiotoxicity marker gene, SEQ ID NO:2438.
XX
KW   Toxicology screening; drug screening; cardiotoxicity, drug-induced;
KW   expression profile; gene expression; myocarditis; heart arrhythmia;
KW   tachycardia; myocardial ischemia; angina; hypertension; hypotension;
KW   dyspnea; cardiogenic shock; cardiovascular disease; gene; ss.
XX
OS   Rattus norvegicus.
XX
PN   WO2004063334-A2.
XX
PD   29-JUL-2004.
XX
PF   08-JAN-2004; 2004WO-US000240.
XX
PR   08-JAN-2003; 2003US-00338044.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Mendrick DL, Porter MW, Johnson KR, Higgs B, Castle A;
PI   Elashoff M;
XX
DR   WPI; 2004-561893/54.
DR   REFSEQ; NM_053835.

```
Db    519 AGCCAGATGCTGATACCATTGCCTATGTGGCATCGGAAGAGAGGCTTTGTGAAAGAATCCA 578
Qy    427 AGGAGGAGACCCCAGGCACAGAGTGGGAGAAGTGGCCCAGCTGTGTGACTTCAACCCTA  486
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    579 AGGAGGAGACCCCAGGCACAGAGTGGGAGAAGTGGCCCAGCTGTGTGACTTCAACCCTA  638
Qy    487 AGGAGCAGCAGCAATGTAAAGACGTGTCCCGCCTGTCCGGTCGTCAT             535
          |||||||||||  |||||||||||||||||||||||  |||||||
Db    639 AGGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTGCGCCTCGGTGGTGCTCAT       687
```

RESULT 9
ABK83950
ID   ABK83950 standard; cDNA; 1051 BP.
XX
AC   ABK83950;
XX
DT   14-AUG-2002 (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #521.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
PN   WO200228999-A2.
XX
PD   11-APR-2002.
XX
PF   03-OCT-2001; 2001WO-US030821.
XX
PR   03-OCT-2000; 2000US-0237189P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
DR   WPI; 2002-435328/46.
XX
PT   Detecting granulocyte activation by detecting differential expression of
PT   genes associated with granulocyte activation, which serves as diagnostic
PT   markers that is useful for monitoring disease states and drug toxicity.
XX
     Claim 1; SEQ ID NO 521; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing the
CC   expression level to an expression level in an unactivated GC, where
CC   differential expression of Gs is indicative of GCA. Also included are
CC   modulating (M2) GA by contacting GC with an agent that alters the
CC   expression of at least one gene in Gs; (2) screening (M3) for an agent
CC   capable of modulating GCA or an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease using the gene expression
CC   profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease, by detecting the level of
CC   expression in a sample of the tissue of gene(s) from Gs, where the level
CC   of expression of the gene is indicative of inflammation; (4) treating
CC   (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease, by contacting a tissue having inflammation with an
CC   agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC   is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC   for screening an agent capable of modulating GCA preferably in an
CC   inflammation in a tissue; M4 is useful for detecting an inflammation
CC   (especially chronic) in a tissue, an allergic response in a subject,
CC   exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC   psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC   cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC   respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC   disease, ulcerative colitis, periodontal disease; also bacterial
CC   infection, viral infection, parasitic infection, protozoal infection,
CC   fungal infection and M5 is useful for treating one of the above
CC   conditions. The present sequence represents a gene differentially
CC   expressed in granulocytes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1051 BP; 209 A; 315 C; 361 G; 166 T; 0 U; 0 Other;

Query Match   84.4%; Score 451.8; DB 6; Length 1051;
Best Local Similarity   90.3%; Pred. No. 2.1e-110;
Matches 483; Conservative   0; Mismatches 52; Indels   0; Gaps   0;

```
Qy     1 CGGCCGCCTTCCTGGCCCAGCAGGAGGAGGAGGATTGCTGGCATCGAGAATGACTCGGGT  60
         ||||||||||||||||||||||||||  |||||||||| ||||| |||  |||  ||||
Db   210 CGGCCGCCTTCCTGGCCCAGCAGGAGAGCGGAGATTGCAGGCATAGAGAACGACGAGGGCT 269

Qy    61 TCGGGGCACCTGCCGCCAGCCAGGTGGCCTCTGCCGCCCGGACTCGCGAGCGGGGGGTG  120
         |||||||||||||||||||||| ||| ||   ||||||||  |||||||||| ||| ||
Db   270 TCGGGGCACCTGCCGCCAGCCATGCGGCCCCCCGGCCAGCCGGCCCCACGAGTGGGGCTG  329

Qy   121 GTTCGGAGGACATGGGGACTACAGTCAATGGGGATGTGTTCACGAGGGCTAACGGGCCTG 180
         |||| |||||||||||||| | |||||||||||||||||| ||| || |||  |||||||
Db   330 GTTCTGAGGACATGGGGACCACAGTCAATGGGGATGTGTTTCAGGAGGCCAACGGTCCTG 389

Qy   181 CCGATGGCTACGTGCGATTGCCCAGGCGACAGGTTGACTCAGGAGCCTGAGAGCATCC   240
         | |||||||||| | ||||||||||||  ||||  |||  ||||||||||||||||||
Db   390 CTGATGGCTACGGACGCCATTGCCCAGGCTGACAGGCTGACCCCAGGAGCCTGAGAGCATCC 449

Qy   241 GCAAGTGGAGAGGAGGAGAACGGCTGAGGAGTTGAATGCTGCCTCGAAGGTGA        300
         ||||||||  ||||||||| ||| ||| || ||  | |||||| || |||| |
Db   450 GCAAGTGGGCAGAGGAGGAGGAGAACGCGCTGCAAGAGTCTGGATGCTGCATCTAAGGTCA 509

Qy   301 CCGAACAGGAGTGCGGCGGAGGAAGCGCCAAAAAAGACCTGGACGAGTGGAACCAGGCCAAA 360
         | ||||||||  ||| ||||||||||||||  || |||||||| ||||||||||||| |||
Db   510 CGGAACAGGAATGCGCGGAGGAAGCGCCAAGAAGGACCTGGACGAGTGGAACCAGCGCCAGA 569

Qy   361 GTGAACAGTTGAAGAAGAACAAGATCAACAACAGGGCATCGGAAGAGGGCTTTGTGAAAG 420
         ||||||||| |||||||||||||||||||||| ||||| | ||||| |||||| |||| |
Db   570 GTGAACAGTAGAGAAGAACAAGATCAACAACAACCCGGCATCCGGAGGAGGCTTTCGTGAAGG 629

Qy   421 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGTGGCCCAGCTGTGTGACTTCA  480
         |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   630 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGTGGCCCAGCTATGTGACTTCA  689

Qy   481 ACCCTAAGAGCAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTGTCCGGTCGTCAT    535
         ||||  ||||||||||||||||| ||||||| |||||||||||||  |   ||| ||
Db   690 ACCCCAAGAGCAGCAGCAAGCAAGTGCAAAGATGTGTCCCGCCTGCGCCTCGGTGGTGCTCAT 744
```

RESULT 10
ADJ62744
ID   ADJ62744 standard; cDNA; 1051 BP.
XX
AC   ADJ62744;
XX
DT   06-MAY-2004 (first entry)
XX
DE   Human cDNA differentially expressed in response to docetaxel #14.
XX
KW   array; docetaxel; docetaxel resistance; cancer; human; ss;
KW   differential expression.
XX
OS   Homo sapiens.
XX

PN   US2004018527-A1.
XX
PD   29-JAN-2004.
XX
PF   16-MAY-2003; 2003US-00439703.
XX
PR   17-MAY-2002; 2002US-0381141P.
XX
PA   (CHAN/) CHANG J C.
PA   (OCON/) O'CONNELL P.
XX
PI   Chang JC, O'connell P;
XX
DR   WPI; 2004-224389/21.
XX
PT   Array useful for screening patient for resistance to docetaxel comprises
PT   complementary nucleic acid probes attached to solid surface.
XX
PS   Claim 21; SEQ ID NO 14; 169pp; English.
XX
CC   The invention relates to an array for screening a patient for resistance
CC   to docetaxel comprising complementary nucleic acid probes attached to a
CC   solid surface for at least 10 of the nucleic acids chosen from 91 fully
CC   defined sequences as given in the specification. The array is useful for
CC   screening a patient for resistance to docetaxel. The array is also useful
CC   for monitoring a cancer patient receiving docetaxel therapy. The present
CC   sequence represents a human cDNA differentially expressed in response to
CC   docetaxel.
XX
SQ   Sequence 1051 BP; 209 A; 315 C; 361 G; 166 T; 0 U; 0 Other;

Query Match   84.4%; Score 451.8; DB 12; Length 1051;
Best Local Similarity   90.3%; Pred. No. 2.1e-110;
  Matches 483; Conservative   0; Mismatches 52; Indels   0; Gaps   0;

Qy   1 CGGCCGCCTTCCTGGCCCAGCAGCAGGAGGCGAGATTGCTCGGCATCGAGAATGACTCGGTT 60
              ||||||||||||||||||||||||||||||||||||| ||| ||||||||| |||| |||
Db   210 CGGCCGCCTTCCTGCCCCAGCAGCAGGAGGCGAGATTGCAGGCATAGAGAACGACGAGGGCT 269

Qy   61 TCGGGGCACCTGCGCGCCAGCCAGGTGGCCTCTGCGCAGCCGGGTG 120
              |||||||||| |||| |||||||| ||||| |||||
Db   270 TCGGGGCACCTGCGCGCCAGCCAGGTGGGCAGCCCCGAGCGGGGCTG 329

Qy   121 GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG 180
              ||||| |||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   330 GTTCTGAGGACATGGGACCACCAGTCAATGGAGATGTGTTTCAGGAGGCCAACGGTCCTG 389

Qy   181 CCGATGGCTACGCTGCGATTGCCCAGGCCGGACAGGTTGACTCAGGAGCCTGAGAGCATCC 240
              ||||||||||||||||||||| ||||| ||| ||||| |||||| |||||| ||||||||
Db   390 CTGATGGCTACGCAGCCATTGCCCAGGCTGACAGGCTGACACCCAGGAGCCTGAGAGCATCC 449

Qy   241 GCAAGTGGAGAGAGGAGCAGAAGAAAGGCTGCAAGAGCTGGATGCTGCCTCGAAGGTGA 300
              |||||||| ||||||||||||||||||||||||||||||| ||||| ||| |||||||
Db   450 GCAAGTGGCCGAGAGGAGCAGAAGAAAACGGCTGCAAGAGCTGGATGCTGCTCAAGGTCA 509

Qy   301 CCGAACAGGAGTGGCCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 360
              ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   510 CGGAACAGGAATGGCCGGGAGAAGGCCAAGAAGCCAAGAAGGCCAAGAAGGAGTGGAACCAGCGCCAGA 569

Qy   361 GTGAACAGGTTGAGAGAGAACAAGATCAACAACAGCGGCATCGGAAGAGGCTTTTGTGAAAG 420
              |||||||| ||||| |||||||||||||||||||||| |||| ||||||||||| |||
Db   570 GTGAACAAGTAGAGAGAGAACAAGATCAACAACACCGGCATCCGAGGAGGCTTTCGTGAAGG 629

Qy   421 AATCCAAGGAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCA 480
              |||||||||||||||||||||||||||||||||| |||||| ||||| ||| ||||||||
Db   630 AATCCAAGGAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTATGTGACTTCA 689

Qy   481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCGCTTCGGGTCGGTGCTCAT 535
              |||| ||||||||| |||||| ||||| |||||||||||||||||||||||||||||
Db   690 ACCCCAAGAGCAGCAAGCAGTGTCGCCCGCTTCCCGCGCTTCCGGGTCGGTGCTCAT 744

RESULT 11

ADX07445
ID   ADX07445 standard; DNA; 1051 BP.
XX
AC   ADX07445;
XX
DT   21-APR-2005   (first entry)
XX
DE   Cyclin-dependent kinase modulation biomarker DNA SEQ ID NO 2011.
XX
KW   cytostatic; cyclin-dependent kinase; cdk; biomarker; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2005012875-A2.
XX
PD   10-FEB-2005.
XX
PF   29-JUL-2004; 2004WO-US024424.
XX
PR   29-JUL-2003; 2003US-0490890P.
XX
PA   (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI   Li M, Rupnow BA, Webster KR, Jackson DG, Wong TW;
XX
DR   WPI; 2005-163068/17.
DR   P-PSDB; ADX07446.
XX
PT   Biomarkers useful for predicting or determining the response of a mammal
PT   to a cancer treatment comprising administration of a modulator of cyclin-
PT   dependent kinase activity.
XX
PS   Claim 5; SEQ ID NO 2010; 141pp; English.
XX
CC   This invention describes a novel method of predicting or determining
CC   whether a mammal will respond or is responding to an anti-cancer agent
CC   that modulates cyclin-dependent kinase (cdk) activity. The method
CC   comprises measuring the level of one or more biomarkers selected from
CC   2774 biomarkers given in the specification (nucleotide sequence SEQ ID
CC   NO:1246 (Genbank EST W28729) is especially preferred). The method of the
CC   invention is utilized in a kit for determining or predicting whether
CC   patient would be susceptible or resistant to treatment by an agent
CC   modulating cdk activity. The invention also describes a method for
CC   utilizing individualized genetic profiles for treating diseases and
CC   disorders based on patient's response and molecular level, specialized
CC   microarrays comprising the biomarkers described, antibodies directed
CC   against the biomarkers and a cell culture model to identify biomarkers.
CC   The cdk modulator is preferably N-5-[(5-(1,1-Dimethylethyl)-2-
CC   oxazolyl]methylthio]-2-thiazolyl-4-piperidine carboxamide, 0.5-L-
CC   tartaric acid salt. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct sequences. This
CC   sequence encodes a biomarker used in the method of the invention.
XX
SQ   Sequence 1051 BP; 209 A; 315 C; 361 G; 166 T; 0 U; 0 Other;

Query Match   84.4%; Score 451.8; DB 14; Length 1051;
Best Local Similarity   90.3%; Pred. No. 2.1e-110;
  Matches 483; Conservative   0; Mismatches 52; Indels   0; Gaps   0;

Qy   1 CGGCCGCCTTCCTGGCCCAGCAGCAGGAGGCGAGATTGCTCGGCATCGAGAATGACTCGGTT 60
              ||||||||||||||||||||||||||||||||||||| ||| ||||||||| |||| |||
Db   210 CGGCCGCCTTCCTGCCCCAGCAGCAGGAGGCGAGATTGCAGGCATAGAGAACGACGAGGGCT 269

Qy   61 TCGGGGCACCTGCGCGCCAGCCAGGTGGCCTCTGCGCAGCCGGGGTG 120
              |||||||||| |||| |||||||| ||||| |||||
Db   270 TCGGGGCACCTGCGCGCCAGCCAGGTGGGCAGCCCCGAGCGGGGGCTG 329

Qy   121 GTTCGGAGGACATGGGACTACAGTCAATGTGTTTCAGGAGGCTAACGGGCCTG 180
              ||||| |||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   330 GTTCTGAGGACATGGGACCACCAGTCAATGTGTTTTCAGGAGGCCAACGGTCCTG 389

Qy   181 CCGATGCTACGGCTGCGATTGCCCAGGCCGGACAGGTTGACTCAGGAGCCTGAGAGCATCC 240

```
Db    390 CTGATGGCTACGCAGCCATTGCCCAGGCTGACAGGCTGAGAGCCTGAGAGCATCC 449
          |||||||| || |||||||| |||||| |||| ||||||||||||||||||||||
Qy    241 GCAAGTGGCGGAGGAGCAGAGAAAAGGCTGGATGCTGCGAGGAGTTGGATGCTGCCTCGAAGGTGA 300
          |||||||||||||||||||||||||||||||| ||
Db    450 GCAAGTGGCGGAGGAGCAGAGAAACGGCTGCAAGAGCTGCGATGCTGCATCTAAGGTCA 509
Qy    301 CCGAACAGGAGTGCGCGGGAGAAGGCCAAAAAAGACCTGGAACCAGCGCCAAA 360
          |||||||||||||||||||||||||||| || |||||||||||||||||||||
Db    510 CGGAACAGGAATGCGGGGAGAAGGCCAAGAAGACCTGGAGGAGTGGAACCAGCGCCAGA 569
Qy    361 GTGAACACAGGTTGAGAAGAACAAGATCAACAACACAGGGCTCGGAAGAGGCTTTTGTGAAAG 420
          ||||| |||| ||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    570 GTGAACAAGTAGAGAAGAACAAGATCAACAACACCGGGCATCCGAGGAGGCTTTCGTGAAGG 629
Qy    421 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCA 480
          |||||||||||||||||||| ||||||||||| ||||||||||||||||||||||||||||
Db    630 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTATGTGACTTCA 689
Qy    481 ACCCTAAGAGACAGCAAGCAATGTAAAGACGTGTCCCGCTCGGTGCTCAT 535
          |||| ||| |||||||||||||||||||||||||||||||||||||||||
Db    690 ACCCCAGAGAGCAGCAAGCAGTGCAAAGATGTGTCCCGCTCGGTGCTCAT 744
```

```
RESULT 12
ADX07447
ID   ADX07447 standard; DNA; 1134 BP.
XX
AC   ADX07447;
XX
DT   21-APR-2005  (first entry)
XX
DE   Cyclin-dependent kinase modulation biomarker DNA SEQ ID NO 2012.
XX
KW   cytostatic; cyclin-dependent kinase; cdk; biomarker; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2005012875-A2.
XX
PD   10-FEB-2005.
XX
PF   29-JUL-2004; 2004WO-US024424.
XX
PR   29-JUL-2003; 2003US-0490890P.
XX
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI   Li M,  Rupnow BA,  Webster KR,  Jackson DG,  Wong TW;
XX
PT   Biomarkers useful for predicting or determining the response of a mammal
PT   to a cancer treatment comprising administration of a modulator of cyclin-
PT   dependent kinase activity.
XX
PS   Claim 5; SEQ ID NO 2012; 141pp; English.
XX
CC   This invention describes a novel method of predicting or determining
CC   whether a mammal will respond or is responding to an anti-cancer agent
CC   that modulates cyclin-dependent kinase (cdk) activity. The method
CC   comprises measuring the level of one or more biomarkers selected from
CC   2774 biomarkers given in the specification (nucleotide sequence SEQ ID
CC   NO:1246 (Genbank EST W28729) is especially preferred). The method of the
CC   invention is utilized in a kit for determining or predicting whether
CC   patient would be susceptible or resistant to treatment by an agent
CC   modulating cdk activity. The invention also describes a method for
CC   utilizing individualized genetic profiles for treating diseases and
CC   disorders based on patient's response and molecular level, specialized
CC   microarrays comprising the biomarkers described, antibodies directed
CC   against the biomarkers and a cell culture model to identify biomarkers.
CC   The cdk modulator is preferably N-5-[(5-(1,1-Dimethylethyl)-2-
```

```
CC   oxazolyl]methyl]thio]-2-thiazolyl-4-piperidine carboxamide, 0.5-L-
CC   tartaric acid salt. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published/pct sequences. This
XX   sequence encodes a biomarker used in the method of the invention.
SQ   Sequence 1134 BP; 225 A; 338 C; 390 G; 181 T; 0 U; 0 Other;

Query Match            72.5%;  Score 387.8;  DB 14;  Length 1134;
Best Local Similarity   82.0%;  Pred. No. 2.7e-93;
Matches 483;   Conservative    0;  Mismatches   52;  Indels   54;  Gaps    1;

Qy      1 CGGCCGCCTTCCTGGCCACGCAGGAGGGACGATTGCTGCGATCAGCGAGATGACTCGGGTT  60
          |||||||||||||||||| ||||| ||||||||| ||||||||||||||||||||||||
Db    243 CTGATGGCTACGCAGCCATTGCCCAGGCTGACAGGCTGACAGGCTGAGAGCATCC  302
Qy     61 TCGGGGCACCTGCCGCCAGCCAGTGGCCTTCGCGGCAGCCCGGACTCGCAGCGCGGGGGTG 120
          ||||||||||||||||||||||||||||||| ||  |||||||||||||||||||||||||
Db    303 TCGGGGCACCTGCCGCCAGCCATGGCCTTCGCGGCAGCCCGGGACCCCACGAGTGGGGCTG 362
Qy    121 GTTCGGAGGACATGGGACTACAGTCAGTCAATGGATGTGTTTCAGGAGGGTAACGGCCTG 180
          ||||||||||||||||| ||||||||| ||||||||||||||| ||||||||||||||||
Db    363 GTTCTGAGGACATGGGACCACAGTCAATGGAGATGTGTTTCAGGAGGCCAACCGTCCTG 422
Qy    181 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGCAGGTTGACTCAGGAGCCTGAGAGCATCC 240
          ||||||||||||| |||| ||||||||| ||||||||||||| |||||||||||||||||||
Db    423 CTGATGGCTACGCAGCCATTGCCCAGGCTGACAGGCTGACAGGCTGAGAGCTGAGAGCATCC 482
Qy    241 GCAAGTGAGAGGAGGAGCAGAGAAAAGGCTGGATGCTGCTGCCTCGAAGGTGA 300
          |||||||||||||||||||||||||||| ||||||||||| |||||||||||||
Db    483 GCAAGTGGCGGAGGAGCAGAGGAGCAGGAAACGGCTGCTCGATGCTGGATGCTGCATCTAAGGTCA 542
Qy    301 CCGAACAGGAGTGCGCGGGAGAAGGCCAAAAAAAGACCTGGAACCAGCGCCAAA 360
          ||||||||||||||| |||||||||||| ||  |||||||||||||||||||||
Db    543 CGGAACAGGAATGCGGGGAGAAGGCCAAGAAGACCTGGAGGAGTGGAACCAGCGCCAGA 602
Qy    361 GTGAACAGTTGAGAGAACAAGAACAAGATCAACAACAG--------------------- 394
          ||||||| ||||  ||| ||||||||||||||||||||
Db    603 GTGAACAAGTAGAGAAGAACAAGATCAACAACACCGGCGATCGCTGACAAAGCATTCTACCAGC 662
Qy    395 ------------------------GGCATCGGAAGAGGCTTTTGTGAAAGAATCCA 426
                                   ||||||||||||||||||| |||||||||||||
Db    663 AGCCAGATGCTGATATCATCGGCTACGTGGCATCCGAGGAGGCTTTCGTGAAGGAATCCA 722
Qy    427 AGGAGGAGCACACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTA 486
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    723 AGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTATGTGACTTCAACCCCA 782
Qy    487 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCTCGGTGCTCAT 535
          ||||||||||||| ||| |||| ||||||||||||||||||||
Db    783 AGAGCAGCAAGCAGTGCAAAGATGTGTCCCGCTCGGTGCTCAT 831
```

```
RESULT 13
ABK34444/c
ID   ABK34444 standard; cDNA; 894 BP.
XX
AC   ABK34444;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Human cDNA for novel secreted protein, SEQ ID 213.
XX
KW   Human; ss; gene; secreted protein; immune deficiency; viral infection;
KW   bacterial infection; fungal infection; autoimmune disorder; burn;
KW   rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;
KW   diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
KW   Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
KW   coagulation disorder; haemophilia; inflammatory disorder; ulcer;
KW   tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
KW   lymphoid cell deficiency.
XX
OS   Homo sapiens.
```

XX
PN   WO200177290-A2.
XX
PD   18-OCT-2001.
XX
PF   29-MAR-2001; 2001WO-US010295.
XX
PR   06-APR-2000; 2000US-0194941P.
XX
PA   (GEMY) GENETICS INST INC.
XX
PI   Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI   Gulukota K, Graham JR;
XX
DR   WPI; 2002-179323/23.
XX
PT   Six hundred and twenty five polynucleotides derived from a variety of
PT   human tissue sources which encode secreted proteins, useful for treating
PT   immune deficiencies and disorders such as autoimmune disorders.
XX
PS   Claim 1; Page 129; 339pp; English.
XX
CC   The invention relates to 625 polynucleotides which have been derived from
CC   a variety of human tissue sources and which encode novel secreted
CC   proteins, their complements and sequences that hybridise to them. Also
CC   included are a vector comprising the polynucleotide, a host cell
CC   transformed with the vector, the proteins encoded by the polynucleotides,
CC   antibodies that bind to the proteins and identification of modulators of
CC   the proteins or the expression of the polynucleotide. The polynucleotides
CC   can be used as probes for the identification and isolation of full length
CC   cDNA and genomic DNA. The polynucleotides and proteins can also be used
CC   as nutritional supplements. The protein is useful in the treatment of
CC   various immune deficiencies and disorders such as viral infections,
CC   bacterial infections, fungal infections, autoimmune disorders (e.g.
CC   rheumatoid arthritis, multiple sclerosis, autoimmune thyroiditis and
CC   diabetes) and allergic reactions and conditions (e.g. asthma). They are
CC   also useful for treating neurodegenerative diseases (e.g. Alzheimer's
CC   disease, Parkinson's disease), liver fibrosis, coagulation disorders
CC   (e.g. haemophilia), inflammatory disorders (e.g. Crohn's disease) and
CC   tumours. They are also useful for tissue regeneration, for wound healing
CC   and in the treatment of burns, incisions and ulcers. The proteins are
CC   also useful for regulating haematopoiesis, for treating myeloid or
CC   lymphoid cell deficiencies. The present sequence is one of the 625 cDNA
CC   sequences encoding a secreted protein
XX
SQ   Sequence 894 BP; 154 A; 280 C; 254 G; 206 T; 0 U; 0 Other;

Query Match   71.2%; Score 380.8; DB 6; Length 894;
Best Local Similarity 81.8%; Pred. No. 1.8e-91;
Matches 476; Conservative   0; Mismatches 52; Indels 54; Gaps   1;

Qy   8 CTTCCTGCCCAGCAGAGCGAGAGCGAGATTGCTCGGCATCGAGAATGACTCGGGTTTCGGGGC 67
            ||||||||||||||||||||||||||||||||| ||| |||||||||||
Db   894 CTTCCTGGCCCAGCAGAGGAGAGCGAGAGATTGCAGGCATAGAGAACCACGAGGGGGCTTCGGGGC 835

Qy   68 ACCTGCCGCGCCAGCCAGTCGCCTTCGCCAGCCCGGACTCGCCAGCCGGCGGGGGGTGTTCGGA 127
            |||||| |||||| |||||| ||| |||||||||||| ||
Db   834 ACCTGCCGCAGCCCATGCGGCCCCCGCGGCCGGCCGCCCCAGCGAGTCGGGGCTGTGTTCTGA 775

Qy   128 GGACATGGGGACTACAGTCAATGAGAGATGTGTTTCAGGAGGCTAACGGGCTGCCGATGG 187
            |||||||||||||||| |||||||||||||||||||| |
Db   774 GGACATGGGGACCACAGTCAATGAGAGATGTGTTTCAGGAGGCCAACGTCCTGCTGATGG 715

Qy   188 CTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCCTCAGGAGCATCCGCAAGTG 247
            ||||| ||||||||||||||||| |||||
Db   714 CTACGCAGCCATTGCCCAGGCCTGACCAGGCCTGACCCAGGAGCCTGAGAGCATCCCGCAAGTG 655

Qy   248 GAGAGAGGAGGCAGAAAAGGCTCCAGGAGTTGGATGCTGCCTCGAAGGTGACCGGAACA 307
            |||||| || |||||||||||
Db   654 GCGAGAGGAGGCAGAGGAAACGGCTGCAAGAGCTGCGATGCTGCATCTAAGGTCACGGAACA 595

Qy   308 GGAGTGGCCGGAGAAGGCCAAAAAGACCTGGACGAGTGGAACCAGCCGCCAAAGTGAACA 367
            |||||||||||| || |||||
Db   594 GGAATGGCGGGAGGAAGGCCAAGAAGACCTGGAGGAGTGGAACCAGCCGCCAGAGTGAACA 535

Qy   368 GGTTGAGAGAACAAGATCAACACAG----------------------------- 394
            |||||| ||| ||||| ||
Db   534 AGTAGAGAGAACAACAAGATCAACACCGGATCGCTGACAAAGCATTCTACCAGCAGCCAGA 475

Qy   395 ----------------------GGCATCGGAAGAGAGGCTTTGTGTAAAGAATCCAAGGAGGA 433
                                  |||||| ||||| ||||||||| |
Db   474 TGCTGATATCATCCGGCTACGTGCCATCGGCATCCGGAGGAGGCTTTCGTGAAGGAATCCAAGGAGGA 415

Qy   434 GACCCCAGGCACAGAGTGGGAGAAGTGGCCCAGCTGTGTGACTTCAACCTAAGACAG 493
            || ||||||| ||||||||| ||||
Db   414 GACCCCAGGCACAGAGTGGGAGAAGTGGCCCAGCTATGTGACTTCAACCCCAGAGCAG 355

Qy   494 CAAGCAATGTAAAGACGTGTCCCGCCTGCTCCGTCCGTGCTCAT 535
            |||||||||
Db   354 CAAGCAGTGCAAAGATGTGTCCCGCCTGCTCCGTCCGTGCTCAT 313

RESULT 14
ABK34409/c
ID   ABK34409 standard; cDNA; 907 BP.
XX
AC   ABK34409;
XX
DT   08-MAY-2002 (first entry)
XX
DE   Human cDNA for novel secreted protein, SEQ ID 178.
XX
KW   Human; sg; gene; secreted protein; immune deficiency; viral infection;
KW   bacterial infection; fungal infection; autoimmune disorder; burn;
KW   rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;
KW   diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
KW   Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
KW   coagulation disorder; haemophilia; inflammatory disorder; ulcer;
KW   tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
KW   lymphoid cell deficiency.
XX
OS   Homo sapiens.
XX
PN   WO200177290-A2.
XX
PD   18-OCT-2001.
XX
PF   29-MAR-2001; 2001WO-US010295.
XX
PR   06-APR-2000; 2000US-0194941P.
XX
PA   (GEMY) GENETICS INST INC.
XX
PI   Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI   Gulukota K, Graham JR;
XX
DR   WPI; 2002-179323/23.
XX
PT   Six hundred and twenty five polynucleotides derived from a variety of
PT   human tissue sources which encode secreted proteins, useful for treating
PT   immune deficiencies and disorders such as autoimmune disorders.
XX
PS   Claim 1; Page 118-119; 339pp; English.
XX
CC   The invention relates to 625 polynucleotides which have been derived from
CC   a variety of human tissue sources and which encode novel secreted
CC   proteins, their complements and sequences that hybridise to them. Also
CC   included are a vector comprising the polynucleotide, a host cell
CC   transformed with the vector, the proteins encoded by the polynucleotides,
CC   antibodies that bind to the proteins and identification of modulators of
CC   the proteins or the expression of the polynucleotide. The polynucleotides
CC   can be used as probes for the identification and isolation of full length
CC   cDNA and genomic DNA. The polynucleotides and proteins can also be used
CC   as nutritional supplements. The protein is useful in the treatment of
CC   various immune deficiencies and disorders such as viral infections,
CC   bacterial infections, fungal infections, autoimmune disorders (e.g.

CC   rheumatoid arthritis, multiple sclerosis, autoimmune thyroiditis and
CC   diabetes) and allergic reactions and conditions (e.g. asthma). They are
CC   also useful for treating neurodegenerative diseases (e.g. Alzheimer's
CC   disease, Parkinson's disease), liver fibrosis, coagulation disorders
CC   (e.g. haemophilia), inflammatory disorders (e.g. Crohn's disease) and
CC   tumours. They are also useful for tissue regeneration, for wound healing
CC   and in the treatment of burns, incisions and ulcers. The proteins are
CC   also useful for regulating haematopoiesis, for treating myeloid or
CC   lymphoid cell deficiencies. The present sequence is one of the 625 cDNA
XX   sequences encoding a secreted protein
SQ   Sequence 907 BP; 162 A; 279 C; 257 G; 209 T; 0 U; 0 Other;

Query Match   70.9%;   Score 379.2; DB 6; Length 907;
Best Local Similarity   81.6%;   Pred. No. 4.9e-91;
  Matches   475; Conservative   0; Mismatches   53; Indels   54; Gaps   1;

Qy   8 CTTCCTGCCCAGCAGCAGGAGCGGAGATTGCTGGCATCGAGAATGACTCGGGTTTCGGGGC 67
            ||||||||||||||||| |||||| |||||||||||||| ||||||||||||||| |||||
Db   907 CTTCCTGCCCAGCAGCAGGAGCGGAGATTGCAGGCATAGAGAACGACCGAGGGCTTCGGGGC 848

Qy   68 ACCTGCGCCGCCAGCTGGCCTCTGCGCAGCCCGGACTCCGCAGCGGGCGGGGGTGGTTCGGA 127
            | |||||| ||||||| |||||| |||||||||||||||||||||||| |||| ||||||||
Db   847 ATCTGCCGGCAGCAGCTGCGGCCCCCCGCGCAGCCCGGCGCCCCCACGAGTGGGGGTGGTTCTGA 788

Qy   128 GGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCCTAACGGCCTGCCGATGG 187
            |||||||||| |||||||||||||||||||||||||||||||||||||| |||||||||||
Db   787 GGACATGGGGACCACCAGTCAATGGAGATGTGTTTCAGGAGGCCAACGGTCTCTGCCGATGG 728

Qy   188 CTACGCTGCGATTGCCCAGGCCGACAGGTTGACTCAGGAGGCCTGAGAGCATCCGCAAGTG 247
            |||||||||||| | || |||||||||||||||||||||||||||||| ||||||||||||
Db   727 CTACGCAGCCATTGCCCAGGCCGTGACAGCCTGAGACCCGGGAGGCCTGAGAGCATCCGCAAGTG 668

Qy   248 GAGAGAGGAGCAGAAGAAAAGGCTGCAGGAGTTGGATGCTGCCTGCCGAAGGTGACCGAACA 307
            |||||||||| ||||||||| |||||| |||||||||||||| || |||| |||||||||||
Db   667 GCGAGAGGAGCAGAGGAAAACGGCTGCAAGAGCTGGATGCTGCATCTAAGGTCACGGAACA 608

Qy   308 GGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGACGTGAACCAGCGCCAAAGTGAACA 367
            |||||| | |||| |||||||||||||||||||||||||||||| |||| || |||||||
Db   607 GGAATGGCGGGAGAAGGCCACCAGAGGACCTGGAGAGTGGAACCAGCGCCAGAGTGAACA 548

Qy   368 GGTTGAAGAAGAAGAACAAGATCAACAACAG----------------------- 394
            |||||||||||||||||||||||||||||||
Db   547 AGTAGAGAAGAAGAACAAGATCAACAACCGATCCTGACAAAGCATTCTACCAGCAGCCAGA 488

Qy   395 ---GGCATCGGAAGAGGCTTTTGTGAAAGAATCCAAGGAGGA 433
                                                       |||||||||||||||||||||||||||||||
Db   487 TGCTGATATCATCGGCTACGTGGCATCCGAGGAGGCTTTCGGTGAAGAATCCAAGGAGGA 428

Qy   434 GACCCCAGGCACAGAGTGGGGAGAGGTGGCCCAGCTGTGCCCAGCTTCAACCTTAAGCAGCAG 493
            |||||| ||||||||||||| |||||| ||||||||| | ||||||||||||||| ||||||
Db   427 GACCCCACCAGCAGAGTGGGGCGGGCTGGTGGCCCAGCTATGTGACTTCAACCCCCAAGAGCAG 368

Qy   494 CAAGCAATGTAAAGACGTGTCCCGCCTCGGTGCTCAT 535
            ||||||||||||||||||||||||||||||||||||
Db   367 CAAGCAGTGCAAAGATGTGTCCCGCCTCGCGTGCTCAT 326

RESULT 15
ACH32890
ID   ACH32890 standard; cDNA; 490 BP.
XX
AC   ACH32890;
XX
DT   13-OCT-2003 (first entry)
XX
DE   Human endothelial cell cDNA #1023.
XX
KW   Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW   genome mapping; biodiversity; genetic disorder.
XX
OS   Homo sapiens.
XX

PN   US2003073623-A1.
XX
PD   17-APR-2003.
XX
PF   30-JUL-2001; 2001US-00918995.
XX
PR   30-JUL-2001; 2001US-00918995.
XX
PA   (DRMA/) DRMANAC R T.
PA   (LABA/) LABAT I.
PA   (STAC/) STACHE-CRAIN B.
PA   (DICK/) DICKSON M C.
PA   (JONE/) JONES L W.
XX
PI   Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
DR   WPI; 2003-615964/58.
XX
PT   New polynucleotide sequences obtained from various cDNA libraries, useful
PT   as hybridization probes, as oligomers for PCR, for chromosome and gene
PT   mapping, in the recombinant production of protein, or in generating
PT   antisense DNA or RNA.
XX
PS   Claim 1; SEQ ID NO 20102; 44pp; English.
XX
CC   The invention relates to an isolated polynucleotide comprising any one of
CC   38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC   determined by the technique of SBH (sequencing by hybridisation). Also
CC   included is a purified polypeptide comprising a sequence corresponding to
CC   a reading frame of the novel polynucleotide. The nucleic acid sequences
CC   are useful in diagnostics as expressed sequence tags (EST) for
CC   identifying expressed genes or for physical mapping of the human genome,
CC   in forensics, in assessing biodiversities, or in identifying mutations
CC   responsible for genetic disorders and other traits. The nucleotide
CC   sequences are also useful as hybridisation probes, as oligomers for PCR,
CC   for chromosome and gene mapping, in the recombinant production of
CC   protein, or in generating antisense DNA or RNA. The purified polypeptide
CC   is useful for generating antibodies specific for it. The present sequence
CC   is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ   Sequence 490 BP; 125 A; 117 C; 165 G; 75 T; 0 U; 8 Other;

Query Match   68.1%;   Score 364.4; DB 9; Length 490;
Best Local Similarity   89.3%;   Pred. No. 3.6e-87;
  Matches   392; Conservative   0; Mismatches   47; Indels   0; Gaps   0;

Qy   97 AGCCCGGACTCGCGAGCCGGGGGTGGTTCGGAGACATGGGACTACAGTCAATGGAGATG 156
            ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   44 AGCCGGGCCCCACGAGTGGGGTGGTTCTGAGGACATGGGACCACCAGTCAATGGAGATG 103

Qy   157 TGTTTCAGGAGGCCTAACGGCCTGCCGATGGCTACGCTGCGATTGCCCAGGCGGGACAGGT 216
            ||||||||||||| ||| || ||||||||||||||||| || ||| || ||||||||||||
Db   104 TGTTTCAGGAGGCCAACGGTCTCTGCCGATGGCTACGCAGCCATTGCCCAGGCCGTGACAGGT 163

Qy   217 TGACTCAGGAGGCCTGAGAGCATCCGCAAGTGGAGAGAGGAGCAGAAGAAAAGGCTGCAGG 276
            |||||||||||||||||||||||||||||||||| |||||||||||| |||||| ||||||
Db   164 TGACTCAGGAGGCCTGAGAGCATCCGCAAGTGGCGAGAGGAGCAGAGGAAAACGGCTGCAAG 223

Qy   277 AGTTGGATGCTGCCTGCCGAAGGTGACCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACC 336
            |||||||||| || ||| ||| ||||||||||||||| | |||| ||||||||||
Db   224 AGCTGGATGCTGCATCTAAGGTCACGGAACAGGAATGGCGGGAGAAGGCCACC 283

Qy   337 TGGAGGAGTGGAACCAGCGCCAAAGTGAACAGTTGAGAAGAAGAACAAGATCAACAACAGGG 396
            |||||| |||| |||||||| |||| || |||| |||||||||||||||||||||||||
Db   284 TGGAGGAGTGGAACCAGCGCCAGAGTGAACAAGTAGAGAAGAAGAACAAGATCAACAACCGGG 343

Qy   397 CATCGGAAGAGGCTTTGTGAAAGAATCCAAGGAGGACCCCAGGCACAGAGTGGGGAGA 456
            |||||||||||||| ||||||||||||||||||||| ||||| ||||||||||||| |
Db   344 CATCCGAGGAGGCTTTCGGTGAAGAATCCAAGGAGGAGACCCCACCAGCAGAGTGGGGAGA 403

us-10-664-705-145.rng

```
Qy     457 AGGTGGCCAGCTGTGTGACTTCAACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCC 516
           ||||||||||||| ||||||||||||| |||||||| ||| ||  | ||| |||||||
Db     404 AGGTGGCCCAGCTATGTGACTTCAACCCCAAGAGCAGCAGCANGCAGTGCATAGATGTGTCCCC 463

Qy     517 GCCTGCGCTCGGTGCTCAT 535
           ||||||||||||||||||
Db     464 GCCTGCGCTCGGTGCTCAT 482
```

Search completed: January 5, 2006, 19:41:17
Job time : 492 secs

```
          GenCore version 5.1.6
     Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    January  5, 2006, 19:14:00 ; Search time 3751 Seconds
                                         (without alignments)
                                         6673.178 Million cell updates/sec

Title:         US-10-664-705-145
Perfect score: 535
Sequence:      1 cggccgccttcctggcccag.........cgcctgcctcggtgctcat 535

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched:      41078325 seqs, 2339541228 residues

Total number of hits satisfying chosen parameters:      82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     EST:*
               1: gb_est1:*
               2: gb_est2:*
               3: gb_est3:*
               4: gb_htc:*
               5: gb_est4:*
               6: gb_est5:*
               7: gb_est6:*
               8: gb_est7:*
               9: gb_gss1:*
               10: gb_gss2:*
               11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	c	Score	% Match	Query Length	DB	ID	Description
1		535	100.0	535	1	AA874955	AA874955 UI-R-E0-c
2		535	100.0	682	1	AW914351	AW914351 EST345655
3		506.2	94.6	844	6	CA322501	CA322501 UI-M-FX0-
4		506.2	94.6	1577	2	AK029382	AK029382 Mus muscu
5		503	94.0	651	2	BB614138	BB614138 BB614138
6		499.4	93.3	1026	1	AV129358	AV129358 AV129358
7		498.2	93.1	575	7	CK344101	CK344101 K0835B11-
8		498.2	93.1	612	7	CK344230	CK344230 K0855H11-
9		493.2	92.2	557	2	BF467127	BF467127 UI-M-CG0p
10	c	490.6	91.7	718	3	BI659300	BI659300 603301689
11		472.8	88.4	502	1	AI463714	AI463714 uc45h03.x
12		471.8	88.2	663	5	BY753851	BY753851 BY753851
13	c	471.4	88.1	569	2	BB854705	BB854705 ux32h10.y
14		467.4	87.4	496	2	BF227172	BF227172 uz20f03.x
15		467.2	87.3	476	1	AI104105	AI104105 EST213394
16		466.2	87.1	523	2	BB853395	BB853395 uw31b04.x
17	c	459.8	85.9	496	2	BF720742	BF720742 mab58b02.
18		451.8	84.4	564	1	AW575961	AW575961 UI-HF-BL0
19		451.8	84.4	1071	3	BM563208	BM563208 AGENCOURT
20		451	84.3	451	1	BE098738	BE098738 UI-R-BJ1-
21		447.8	83.7	487	2	BE852625	BE852625 uw26a12.x
22		445.4	83.3	782	3	BP438058	BP438058 BP438058
23		445.4	83.3	838	3	BP440642	BP440642 BP440642
24		443.8	83.0	817	3	BP143717	BP143717 BP143717
25	c	443.2	82.7	796	8	CX091448	CX091448 EHAES5ITR
26		442.2	82.7	825	3	BQ177244	BQ177244 UI-M-ER0-
27		440.8	82.4	678	5	BG470345	BG470345 60253918
28		440.6	82.3	644	5	BY739827	BY739827 BY739827
29	c	440.2	82.3	775	3	BI414098	BI414098 60291348
30		439.2	82.1	659	7	CN409242	CN409242 170004247
31		436.2	81.5	660	5	BY737187	BY737187 BY737187
32		435.4	81.4	759	3	BP438499	BP438499 BP438499
33		434.2	81.2	803	7	CJ030794	CJ030794 CJ030794
34		434.2	81.2	826	8	DN865071	DN865071 nac13f08.
35		434.2	81.2	835	8	DN103678	DN103678 1098884 M
36		432.6	80.9	860	8	DN114523	DN114523 1113900 M
37		432.6	80.9	873	8	DN120155	DN120155 1120145 M
38		432.4	80.8	605	2	BF286069	BF286069 EST450660
39	c	430.2	80.4	516	1	BE646508	BE646508 7e88b03.x
40		429.8	80.3	514	1	AI128183	AI128183 qc43d07.x
41		429.4	80.3	675	7	CK982463	CK982463 4114854 B
42		429.4	80.3	685	7	CK940151	CK940151 4113489 B
43		429.4	80.3	724	7	CK957149	CK957149 4097391 B
44		428.4	80.1	673	7	CK940272	CK940272 4113669 B
45		428.2	80.0	1058	3	BM558124	BM558124 AGENCOURT

ALIGNMENTS

```
RESULT 1
AA874955
LOCUS       AA874955       535 bp    mRNA    linear   EST 05-FEB-1999
DEFINITION  UI-R-E0-ci-g-06-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone
            UI-R-E0-ci-g-06-0-UI 3', similar to gi|203360|gb|M19262|RATCBR3 Rat
            clathrin light chain (LCB3) mRNA, complete cds, mRNA sequence.
ACCESSION   AA874955
VERSION     AA874955.1  GI:4230837
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 535)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
PUBMED      8889548
COMMENT     On Mar 20, 1998 this sequence version replaced gi:2979903.
            Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Oligo-dT track not found. Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.
            Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
            through Research Genetics
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1..535
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-E0-ci-g-06-0-UI"
                     /dev_stage="embryonic"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-E0"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: NotI; Site_2: EcoRI; This library
```

Email: nhlee@tigr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information
Seq primer: M13 Reverse.

```
FEATURES          Location/Qualifiers
     source       1..682
                  /organism="Rattus sp."
                  /mol_type="mRNA"
                  /db_xref="taxon:10118"
                  /clone="RGIAD81"
                  /clone_lib="Normalized rat brain, Bento Soares"
                  /note="Organ: brain; Vector: pT7T3Pac; Site_1: EcoRI;
                  Site_2: NotI"

ORIGIN

Query Match        100.0%;    Score 535;  DB 1;  Length 682;
Best Local Similarity 100.0%;   Pred. No. 4.8e-121;
Matches 535;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy     1 CGGCCGCCTTCCTGGCCCAGCAGGAGAGCGAGATTGCTGCAGAATGACTCGGGTT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CGGCCGCCTTCCTGGCCCAGCAGGAGAGCGAGATTGCTGCAGAATGACTCGGGTT 60

Qy    61 TCGGGGCACCTGCCGCGCCAGTGGCCTCGGCAGCCCGACTCGGCGGGGGGGGGTG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   141 TCGGGGCACCTGCCGCGCCAGTGGCCTCGGCAGCCCGACTCGGCGGGGGGGGGTG 200

Qy   121 GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGCCTG 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   201 GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGCCTG 260

Qy   181 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGGCCTGAGAGCATCC 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   261 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGGCCTGAGAGCATCC 320

Qy   241 GCAAGTGGAGGAGGAGCAGAAGAAGGCTGGCAGGAGTGGATGCTGCCTCGAAGGTGA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321 GCAAGTGGAGGAGGAGCAGAAGAAGGCTGGCAGGAGTGGATGCTGCCTCGAAGGTGA 380

Qy   301 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   381 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 440

Qy   361 GTGAACAGGTTGAGAAGAACAAGATCAACAACAGGGCATCGGAAGAGAGGCTTTGTGAAAG 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   441 GTGAACAGGTTGAGAAGAACAAGATCAACAACAGGGCATCGGAAGAGAGGCTTTGTGAAAG 500

Qy   421 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGTGTGTGACTTCA 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   501 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGTGTGTGACTTCA 560

Qy   481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCGCTCGGTGCTCAT 535
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   621 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCGCTCGGTGCTCAT 675
```

RESULT 3
CA322501
```
LOCUS       CA322501      844 bp    mRNA    linear   EST 09-JUL-2003
DEFINITION  UI-M-FX0-cch-k-13-0-UI.r1 NIH_BMAP_FX0 Mus musculus cDNA clone
            IMAGE:6819686 5', mRNA sequence.
ACCESSION   CA322501
VERSION     CA322501.1  GI:24540599
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 844)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
```

---

consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture."

```
ORIGIN

Query Match        100.0%;    Score 535;  DB 1;  Length 535;
Best Local Similarity 100.0%;   Pred. No. 4.6e-121;
Matches 535;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy     1 CGGCCGCCTTCCTGGCCCAGCAGGAGAGCGAGATTGCTGGCATCGGAATGACTCGGGTT 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CGGCCGCCTTCCTGGCCCAGCAGGAGAGCGAGATTGCTGGCATCGGAATGACTCGGGTT 60

Qy    61 TCGGGGCACCTGCCGCGCCAGTGGCCTCTGCGGACTCGGCAGCCCGACTCGGGGCGGGTG 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TCGGGGCACCTGCCGCGCCAGTGGCCTCTGCGGACTCGGCAGCCCGACTCGGGGCGGGTG 120

Qy   121 GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCCTAACGGGCCTG 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GTTCGGAGGACACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCCTAACGGGCCTG 180

Qy   181 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGGCCTGAGAGCATCC 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGGCCTGAGAGCATCC 240

Qy   241 GCAAGTGGAGGAGGAGCAGAAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCAAGTGGAGGAGGAGCAGAAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 300

Qy   301 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 360

Qy   361 GTGAACAGGTTGAGAAGAACAAGATCAACAACAGGGCATCGGAAGAGAGGCTTTGTGAAAG 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GTGAACAGGTTGAGAAGAACAAGATCAACAACAGGGCATCGGAAGAGAGGCTTTGTGAAAG 420

Qy   421 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGTCCCGCCGCTCGGTGCTCAT 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGTCCCGCCGCTCGGTGCTCAT 480

Qy   481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCGCTCGGTGCTCAT 535
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCGCTCGGTGCTCAT 535
```

RESULT 2
AW914351
```
LOCUS       AW914351      682 bp    mRNA    linear   EST 25-MAY-2000
DEFINITION  EST345655 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
            RGIAD81 5' end, mRNA sequence.
ACCESSION   AW914351
VERSION     AW914351.1  GI:8080027
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 682)
AUTHORS     Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
TITLE       Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
JOURNAL     Unpublished (1998)
COMMENT     Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
```

Fri Jan 6 10:34:24 2006

us-10-664-705-145.rst

```
Db    498 AATCCAAGAGAGGAGGACCCCGGCACAGAGTGGGAGGAAGGTAGCCCAGCTGGGTGTGACTTCA 557
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCGGCCTGCCTCGGTCGGGTGCTCAT 535
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    558 ATCCCAAGAGCAGCAAGCAGTGTAAAGATGTGTCCGGCCTGCCTCGGTCGGGTGCTCAT 612
```

RESULT 4
AKO29382
LOCUS   AKO29382   1577 bp   mRNA   linear   HTC 03-APR-2004
DEFINITION   Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched
             library, clone:4833419E16 product:clathrin, light polypeptide
             (Lcb), full insert sequence.
ACCESSION   AKO29382
VERSION   AKO29382.1   GI:26325351
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS   Carninci,P. and Hayashizaki,Y.
TITLE   High-efficiency full-length cDNA cloning
JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
PUBMED   10349636
REFERENCE   2
AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
             Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
             prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
PUBMED   11042159
REFERENCE   3
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
             Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
             Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
             Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
             Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
             Yoneda,Y., Ishikawa,T., Ozawa,Y., Tanaka,T., Matsuura,S., Kawai,J.,
             Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
             sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
PUBMED   11076861
REFERENCE   4
AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
             FANTOM Consortium.
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
             Group Phase I & II Team.
TITLE   Analysis of the mouse transcriptome based on functional annotation
             of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 1577)
AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
             Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
             Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
             Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
             Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
             Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
             Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
             Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
             Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
             Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,P., Takaku-Akahira,S.,
             Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
             Muramatsu,M. and Hayashizaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
             Physical and Chemical Research (RIKEN), Laboratory for Genome
```

COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
             (BMAP)
FEATURES     Seq primer: pYX-5.
Source       1..844
             /organism="Mus musculus"
             /mol_type="mRNA"
             /strain="C57BL/6"
             /db_xref="taxon:10090"
             /clone="IMAGE:6819686"
             /tissue_type="whole brain"
             /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
             /lab_host="DH10B (T1 phage resistant)"
             /clone_lib="NIH_BMAP_FX0"
             /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
             Site_2: Not I; The library was constructed according
             BonaIdo, Lennon and Soares, Genome Research, 6:791-806,
             1996. Denatured RNA was size fractionated on a 1% agarose
             gel. First strand cDNA synthesis was primed with oligo-dT
             primer containing a Not I site. Double strand cDNA was
             size selected according to mRNA size fraction, ligated
             with EcoR I adaptor, digested with NotI and then cloned
             directionally into pYX-Asc vector. The library tag
             sequence located between the Not I site and the polyA tail
             is AGGAGACAG. This library was created for the University
             Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
             Developing Mouse Nervous System', supported by National
             Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
ORIGIN       program coordinator."

Query Match       94.6%;    Score 506.2;    DB 6;    Length 844;
Best Local Similarity    96.6%;    Pred. No. 6.2e-114;
Matches  517;    Conservative    0;    Mismatches  18;    Indels    0;    Gaps    0;

Qy     1 CGGCGGCCTTCCTGCCCAGCAGGAGAGCGAGATTGCTGCTGGCATCGAGAATGACTCGGGTT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    78 CGGCGGCCGTTCCTGCCCAGCAGGAAAGCGAGATTGCTGCTGGCATCGAGAATGACCCGGGTT 137

Qy    61 TCGGGGCACCTGCCGCCGCCGCCAGGTGGCCTCTGCGCGAGCCCGGACTCGCGAGCCGGGGGTG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   138 TCGGGGCACCTGCCGCCGCCGCCAGGTGGCCTCTGCGCGAGCCCGGACTCGCGAGCCGGGGCTG 197

Qy   121 GTTCGGAGACATGGGGACTACAGTCAATGAGATGTGTTTCAGGAGGGTAACGGCCTG 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   198 GTTCAGAGACATGAGTAGTACACAGTCAATGAGATGTGTTTCAGGAGGCTAACGGCCTG 257

Qy   181 CCGATGGCTACGCTGCGATTGCCCAGGCCGGACAGGTTGACTCAGGAGGCATCC 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   258 CCGATGGCTACGCTGCGATTGCCCAGGCCGGACAGGTTGACTCAGGAGCCCGAGAGCATCC 317

Qy   241 GCAAGTGCGAGAGAGGCAAGAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   318 GGAAGTGCGAGAGAGGCAAGAGAGCAGAGAAAAGGCTGCAGAGTTAGACCGCTGCCTCGAAGGTGA 377

Qy   301 CCGAACAGGAGTGGCGGCGAGAGGCCAAAAAGACCTGGAGGAGTTGAACCAGCGCCAAA 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   378 CCGAACAGGAGTGGGCGGCGAGAGGCCAAGAAAGACCTGGAGGAGTTGGAACCAGCGCCAAA 437

Qy   361 GTGAACAGGTTGAGAAGAACAAGATCAACAACACAGGCATCGGAAGGAGGCTTTTGTGAAAG 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   438 GTGAACAGGTTGAGAAGAACAAGATCAACAACACAGGCATCCGAAGAGAGGCTTTTGTGAAAG 497

Qy   421 AATCCAAGAGAGGAGGACCCCCGGCACAGAGTGGGAGGAAGGTGGCCCAGCTGTGTGACTTCA 480
```

```
Qy   421 AATCCAAGGAGGAGACCCCAGGCCACAGAGTGGGAGAAGGTGGCCAGCTGTGTGACTTCA 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   500 AATCCAAGGAGGAGACCCCGGGCCACAGAGTGGGAGAAGGTGGCCAGCTGTGTGACTTCA 559

Qy   481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCGGCTGCGCTCGGTGCTCAT 535
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   560 ATCCCAAGAGCAGCAAGCAGTGTAAAGATGTGTCCGGCTGCGCTCGGTGCTCAT 614
```

RESULT 5
BB614138
LOCUS   BB614138   651 bp   mRNA   linear   EST 26-OCT-2001
DEFINITION BB614138 RIKEN full-length enriched, 0 day neonate head Mus
            musculus cDNA clone 4833419E16 5', mRNA sequence.
ACCESSION   BB614138
VERSION   BB614138.1 GI:16454597
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 651)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,H., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,T.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
  TITLE   RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
  COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muranatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
            wagi.K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,
            Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
            e mouse tissues.
FEATURES   Location/Qualifiers
     source   1..651
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="4833419E16"
```

---

```
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/
            URL:http://fantom.gsc.riken.jp/.
FEATURES             Location/Qualifiers
     source          1..1577
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM DB:4833419E16"
                     /clone="4833419E16"
                     /tissue_type="head"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="0 day neonate"
     CDS             10..645
                     /note="unnamed protein product; clathrin, light
                     polypeptide (Lcb) (MGD|MGI:1921575, GB|AK009844, evidence:
                     BLASTN, 99%, match=621)
                     /codon_start=1
                     /protein_id="BAC26430.1"
                     /db_xref="GI:26325352"
                     /translation="MAEDFGFPSSSSESGAPEAAEBDPAAAFLAQQESEIAGIENDPGF
                     GAPASQVASAQPGLASGAGSEDMSTTVNGDVFQRANGPADGYAAIAQADRLTQPPES
                     IRKWREEQKKRLQELDAASKVTEQEWREKAKKDLEEWNQRQSEQVEKNKINNRASEEA
                     FVKESKEETPGTEWEKVAQLCDFNPKSSKQCKDVSRLRSVLMSLKQTPLSR"
     polyA_signal    1553..1558
                     /note="putative"
     polyA_site      1577
                     /note="putative"
ORIGIN
Query Match             94.6%;  Score 506.2;  DB 4;  Length 1577;
Best Local Similarity   96.6%;  Pred. No. 6.9e-114;
Matches  517;  Conservative   0;  Mismatches  18;  Indels   0;  Gaps   0;

Qy     1 CGGCGCCTTCTGGCCCAGCAGGAGGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    80 CGGCCGCCTTCCTGGCCCAGCAGGAAGCGAGATTGCTGGCATCGAGAATGACCCGGGTT 139

Qy    61 TCGGGGGACCTGCCGCCCAGCTGGCCTCTGCGCAGCCCGACTCGCGACGCGGGGGTG 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   140 TCGGGGGACCTGCCGCCCAGCTGGCCTCTGCGCAGCCCGACTCGCGACGCGGGGGTG 199

Qy   121 GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCTG 180
           |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   200 GTTCAGGAGGACATGAGTACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCTG 259

Qy   181 CCGATGGCTACGCTGCGATTGCCCAGGCGGCAGGTTGACTCAGGAGGCCTGAGAGCATCC 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   260 CCGATGGCTACGCTGCGATTGCCCAGGCGGCAGGTTGACTCAGGAGGCCCGAGAGCATCC 319

Qy   241 GCAAGTGAGAGTGAGAGAGCAGAGAAGAAAAGGCTGGAGGAGTTGGATTGCTGCCTCGAAGTGA 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   320 GGAAGTGCGAGAGGAGAGAGCAGAGAAGAAAAGGCTGCAGGAGTTAGACGCTGCCTCGAAGTGA 379

Qy   301 CCGAACAGAGGTGGCGGGAGAGGCCAAAAAGACCTTGGAGGAGTGGAACCAGCGCCAAA 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   380 CCGAGACAGAGGTGGCGGGAGAAGACGACCTGGAGGAGTGGAACCAGCGCCAAA 439

Qy   361 GTGAACAGGTTGAGAAGAACAAGATCAACAACAGGGCCATCGGAAGAGGCTTTTGTGAAAG 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   440 GTGAACAGGTTGAGAAGAACAAGATCAACAACAGGGCCATCGGAAGAGGCTTTTGTGAAAG 499
```

```
                    /sex="mixed"
                    /tissue_type="head"
                    /dev_stage="0 day neonate"
                    /lab_host="DH10B"
                    /clone_lib="RIKEN full-length enriched, 0 day neonate
                    head"
                    /note="Site 1: SalI; Site 2: BamHI; cDNA library was
                    prepared and sequenced in Mouse Genome Encyclopedia
                    Project of Genome Exploration Research Group in Riken
                    Genomic Sciences Center and Genome Science Laboratory in
                    RIKEN. Division of Experimental Animal Research in Riken
                    contributed to prepare mouse tissues. 1st strand cDNA was
                    primed with a primer [5'
                    GAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTVN 3'], cDNA was
                    prepared by using trehalose thermo-activated reverse
                    transcriptase and subsequently enriched for full-length by
                    cap-trapper. cDNA went through one round of normalization
                    to Rot = 10.0 and subtraction to Rot = 100.0. Second
                    strand cDNA was prepared with the primer adapter of
                    sequence [5' GAGAGAGAGATTCGAGTTAATTAAATTAATCCCCCCCCCCCC
                    3']. cDNA was cloned into the XhoI and BamHI sites.
                    Vector: a modified pBluescript KS(+) after bulk excision
                    from Lambda FLC I."

ORIGIN

Query Match    94.0%;   Score 503;  DB 2;   Length 651;
Best Local Similarity   96.3%;   Pred. No. 3.6e-113;
Matches 515;   Conservative    0;   Mismatches  20;   Indels    0;   Gaps    0;

Qy     1 CGGCGGCCTTCCTGGCCCAGCAGCAGGAGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    78 CGGCGGCCTTCCTGGCCCAGCAGCAGGAAAGCGAGATTGCTGGCATCGAGAATGACCCGGGTT  137

Qy    61 TCGGGGCACCTGCGCGCCAGCCAGGTGGCCTCGCGCAGCCCGGACTCGCGAGCGGGGGTG  120
         |||||||||||||||||||||||||||||| |||||||||||| ||||||||| ||||
Db   138 TCGGGGCACCTGCGCGCCAGCCAGGTGGCCTCTCGCGCAGCCCGGACTCGCGAGCGGGGCTG  197

Qy   121 GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG  180
         ||| ||||||||| ||||||||||||||||||||||||||||||||||||||| ||||
Db   198 GTTCAGGAGGACATGAGTACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGCCCTG  257

Qy   181 CCGATGGCTACGCTGCGATTGCCCAGGCCGACAGGTTGACTCAGGAGCCTGAGAGCATCC  240
         |||||||||||||||||||||||||||||||||||||| ||||| |||||||||||||
Db   258 CCGATGGCTACGCTGCGATTGCCCAGGCCGACAGGTTGACTCAGGAGCCCGAGAGCATCC  317

Qy   241 GCAAGTGAGAGAGGAGAGCAGAAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA  300
         |||||||||||||||||||||||||||||||||||||| |||| |||||||||||||||
Db   318 GGAAGTGAGAGAGGAGAGCAGAAGAAAAGGCTGCAGGAGTTAGACGCTGCCTCGAAGGTGA  377

Qy   301 CCGAACAGGAGGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA  360
         ||||||||||||||||||||||||||||||||||| ||||||||||||||||||| |||
Db   378 CCGAACAGGAGGTGGCGGGAGAAGGCCAAGAAAGACCTGGAGGAGTGGAACCAGCGCCAAA  437

Qy   361 GTGAACAGGTTGAGAGAGAACAAGATCAACAACAGGGCATCGGAAGAGGCTTTTGTGAAAG  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   438 GTGAACAGGTTGAGAGAGAACAAGATCAACAACAGGGCATCGGAAGAGGCTTTTGTGAAAG  497

Qy   421 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCA  480
         |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   498 AATCCAAGGAGGAGACCCCGGGCACAGAGTGGGAGAAGGTAGCCCAGCTGTGTGACTTCA  557

Qy   481 ACCCTAAGAGCAAGCAATGTAAAGACGTGTCCCGCCTCGGTGCTCAT  535
         |||||||| |||||||||||||||||||||||||||||||||||||
Db   558 AATCCCAGGACAGCAGTGTAAAGATGTGTCCCGCCTCGGTGCTCAT  612

RESULT 6
AV129358
LOCUS       AV129358          1026 bp    mRNA       linear     EST 24-OCT-2001
DEFINITION  AV129358 Mus musculus C57BL/6J 11-day embryo Mus musculus cDNA
            clone 2700067L17, mRNA sequence.
ACCESSION   AV129358
VERSION     AV129358.2   GI:16382589
```

```
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1026)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al 2001)
  JOURNAL   Unpublished (2001)
  COMMENT   On Jul 1, 1999 this sequence version replaced gi:5315593.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res.  10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,
            Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1..1026
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="2700067L17"
                     /sex="mixed"
                     /dev_stage="11-day embryo"
                     /clone_lib="Mus musculus C57BL/6J 11-day embryo"

ORIGIN

Query Match    93.3%;   Score 499.4;  DB 1;   Length 1026;
Best Local Similarity   96.1%;   Pred. No. 3e-112;
Matches 512;   Conservative    0;   Mismatches  21;   Indels    0;   Gaps    0;

Qy     3 GCGGCCTTCCTGGCCCAGCCAGCAGGAGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTTC   62
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   178 GCGGCCTTCCTGGCCCAGCCAGCAGGAAAGCGAGATTGCTGGCATCGAGAATGACCCGGGTTC  237
```

```
Qy    63  GGGGCACCTGCGCGCAGCCAGTGGCCTCTGCGCAGCCCGGACTCGCGCAGCGGGGGTGGT  122
              ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   238  GGGGCACCTGCGCGCAGCCAGTGGCCTCTGCGCAGCCCGCCACTCGCGCAGCGGGGGTGGT  297

Qy   123  TCGGGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGCCTGCC      182
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db   298  TCAGAGACATGAGTACTACAGTGTGGAGATGTGTTTCAGGAGGCTAACGGCCTGCC      357

Qy   183  GATGGCTACGCTGCGATTGCCCAGGCGGCGACAGGTTGACTCAGGAGCATCCGC      242
              ||||   |||||||||||||||||||||||||||||||||||||||||||||
Db   358  GATGGCTACGCTGCGATTGCCCAGGCCGGACAGGTTGACTCAGGAGCATCCGG      417

Qy   243  AAGTGGGAGAGGAGGAGAGCAGAGAAAAAGGCTCAGGAGGTTGGATGCTGCCTCGGAAGGTGACC  302
              ||||||||||||||||||||||||||||||||||||||||||| ||||| |||||||||||
Db   418  AAGTGGGAGAGGAGGAGAGCAGAGAAAAAGGCTGCAGGAGGTTAGACGCTGCCTCGGAAGGTGACC  477

Qy   303  GAACAGGAGTGGGGGGAGAAGGCCAAAAAAGACCTGGAACCAGCGCCAAAGT      362
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   478  GAACAGGAGTGGGGGGAGAAGGCCAGAAAAGACCTGGAACCAGCGCCAAAGT      537

Qy   363  GAACAGGTTGAGAGAAGAACAAGAACAAGATCAACACACAGGCATCGGAAGAGAGGCTTTTGTGAAAGAA  422
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   538  GAACAGGTTGAGAGAAGAACAAGATCAACACACAGGGCATCGGGAGCATCCGGAAGAGAGGCTTTTGTGAAAGAA  597

Qy   423  TCCAAGAGGAGGAGAGACCCCAGGCACAGAGTGGGAGGAAGGTGGCCCAGTGTGTGACTTCAAC  482
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   598  TCCAAGAGGAGGAGACCCCGGGCACAGAGTGGGAGGAAGGTAGCGCCCAGTGTGTGACTTCAAT  657

Qy   483  CCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCTGCCGCTCGGTGCTCAT  535
              || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   658  CCCAAGAGCAGCAAGCAGTGTAAAGATGTGTCCCCGCTGCCGCTCGGTGCTCAT  710
```

```
ORIGIN

Query Match          93.1%;     Score 498.2;  DB 7;     Length 575;
Best Local Similarity 96.6%;     Pred. No. 5.4e-112;
Matches   509;  Conservative   0;  Mismatches   18;  Indels   0;  Gaps   0;

Qy     9  TTCCTGGCCGCCAGCCAGTGGCCGAGATTGCTGGCATCGGCATCGAGAATGACTCGGGTTCGGGCA  68
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  TTCCTGGCCGCCAGCCAGTGGCCGAGATTGCTGGCATCGGCATCGAGAATGACACCCGGGTTCGGGCA  60

Qy    69  CCTGCCGCCAGCCAGTGCCTCTGCGCAGCCCGGACTCGCGCAGCGGGGGGGGTGGTTCGGAG  128
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    61  CCTGCCGCCAGCCAGTGCCTCTGCGCAGCCCGGACTCGCGCAGCGGGGGGGGTGGTTCAGAG  120

Qy   129  GACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGCCTGCCGATGGC  188
              ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   121  GACATGAGTACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGCCTGCCGATGGC  180

Qy   189  TACGCTGCGATTGCCCAGGCGGCGACAGGTTGACTCAGGAGCATCCGCAAGTGG  248
              ||||||||||||||||||||||   |||||||||||||||||||||  ||||||
Db   181  TACGCTGCGATTGCCCAGGCCGGACAGGTTGACTCAGGAGCATCCGGAAGTGG  240

Qy   249  AGAGGAGGAGCAGAGAAAAAGGCTGCAGGAGTTGATGCTGCCTCGGAAGGTGACCGAACAG  308
              ||||||||||||||||||||||||  |||||||| |||||||||||||||||||||||||
Db   241  AGAGGAGGAGCAGAGAAAAAGGCTGCAGGAGTTAGACGCTGCCTCGGAAGGTGACCGAACAG  300

Qy   309  GAGTGGGCGGGAGAAGGCCAAAAAAGACCTGGAACCAGCGCCAAAGTGAACAG  368
              ||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db   301  GAGTGGGCGGGAGAAGGCCAGAAAAGACCTGGAACCAGCGCCAAAGTGAACAG  360

Qy   369  GTTGAGAGAACAAGATCAACACAGGGCATCGGAAGAGGCTTTTGTGAAAGAATCCAAG  428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  GTTGAGAGAACAAGATCAACACAGGGCATCGGAAGAGGCTTTTGTGAAAGAATCCAAG  420

Qy   429  GAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGTGTGTGACTTCAACCCTAAG  488
              ||||||||||||||||||||||||||||  |||||||||||||||||||||||||||||
Db   421  GAGGAGACCCCGGGCACAGAGTGGGAGAAGGTAGCGCCCAGTGTGTGACTTCAATCCCAAG  480

Qy   489  AGCAGCAAGCAATGTAAAGACGTGTCCCGCTGCCGCTCGGTGCTCAT  535
              |||||||||||||  ||||||||||||||| |||||||||||||||||
Db   481  AGCAGCAAGCAGTGTAAAGATGTGTCCCCGCTGCCGCTCGGTGCTCAT  527

RESULT 8
CK344230
LOCUS         CK344230              612 bp    mRNA      linear      EST 22-DEC-2003
```

DEFINITION   K0855H11-3 NIA Mouse 8.5-dpc Whole Embryo cDNA Library (Long) Mus
     musculus cDNA clone NIA:K0855H11 IMAGE:30083710 3', mRNA sequence.
ACCESSION   CK344230
VERSION   CK344230.1   GI:40299843
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
     Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 612)
AUTHORS   Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
TITLE   Construction of long-transcript enriched cDNA libraries from
     submicrogram amounts of total RNAs by a universal PCR amplification
     method
JOURNAL   Genome Res. 11 (9), 1553-1558 (2001)
PUBMED   11544199
COMMENT   Contact: Dawood B. Dudekula
     Laboratory of Genetics
     National Institute on Aging/National Institutes of Health
     333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
     Email: cdna@lgsun.grc.nia.nih.gov
     Plate: K0855   row: H   column: 11
     Seq primer: -21M13 Forward
     High quality sequence stop: 612
     POLYA=Yes.
FEATURES   Location/Qualifiers
     source   1..612
       /organism="Mus musculus"
       /mol_type="mRNA"
       /strain="C57BL/6J"
       /db_xref="niaEST:K0855H11-3"
       /db_xref="taxon:10090"
       /clone="NIA:K0855H11 IMAGE:30083710"
       /tissue_type="whole embryo including extraembryonic
       tissues at 8.5-days postcoitum"
       /dev_stage="8.5-days postcoitum"
       /lab_host="DH10B"
       /clone_lib="NIA Mouse 8.5-dpc Whole Embryo cDNA Library
       (Long)"
       /note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
       NotI; Mouse cDNA project by the Laboratory of Genetics,
       National Institute on Aging (NIA), Intramural Research
       Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
       a long-transcript enriched cDNA library (Ref. Genome Res.
       11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
       extracted from a pool of 13 embryos at 8.5-days
       postcoitum. Double-stranded cDNAs were synthesized with an
       Oligo(dT) primer [Invitrogen:
       5'-PGACTAGTTCTAGATCGCGAGCGGCCGCCCCTTTTTTTTTTTTTT-3'] from
       9.1 ug of total RNA, treated with T4 DNA polymerase, and
       purified by ethanol-precipitation. The cDNAs were ligated
       to Lone-linker LL-Sal4, purified by phenol/chloroform, and
       separated from free linkers by Centricon 100. Then, the
       cDNAs were amplified by long-range high fidelity PCR using
       Ex Taq polymerase (Takara) with a primer Sal4-S. The
       products were purified by phenol/chloroform and Centricon
       100. The cDNAs were digested with SalI and NotI enzymes
       and cloned into SalI/NotI site of pSPORT1 plasmid vector.
       The DH10B E. coli host was transformed with the ligation
       mixture by the standard chemical method. The average
       insert size is about 2.5 kb. The library was constructed
       by Yulan Piao (NIA)."

ORIGIN

Query Match   93.1%;   Score 498.2;   DB 7;   Length 612;
Best Local Similarity   96.6%;
Matches   509;   Conservative   0;   Mismatches   18;   Indels   0;   Gaps   0;

Qy   9 TTCCTGGCCCAGCCAGGAGGCGAGATTGCTGGCATCGAGAATGACTCGGGTTTCGGGGCA 68
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 TTCCTGGCCCAGCCAGGAAAGCGAGATTGCTGGCATCGAGAATGACCCGGGTTTCGGGGCA 60

Qy   69 CCTGCGCCAGCCAGGTGGCCTCTGCGCAGCCCGGACTCGCGAGCGGGGGTGGTTCGGAG 128
     ||||||||||||||||||||||||||||||||||||| |||||||||| |||||||||||
Db   61 CCTGCGCCAGCCAGGTGGCCTCTGCGCAGCCCGGACTCGCGAGCGGGGGCGGTTCAGAG 120

Qy   129 GACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTGCCGATGGC 188
     |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GACATGAGTACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTGCCGATGGC 180

Qy   189 TACGGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCCTGAGAGCCATCCGCAAGTGG 248
     |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   181 TACGGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCCCGAGAGCCATCCGGAAGTGG 240

Qy   249 AGAGGAGGAGCAGAAGCTGCAGGAGTTGATGCTGCCTGCAGGAGTGACCGAACAG 308
     |||||||||||||||||||||||||||||| ||| |||||||||||||||||||
Db   241 AGAGGAGGAGCAGAAGCTGCAGGAGTTAGACGCTGCCTGCTCCAGGAGTGACCGAACAG 300

Qy   309 GAGTTGCGGGAGAAGGCCAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAAGTGAACAG 368
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GAGTTGCGGGAGAAGGCCAAGAAGACCCGGAGGAGTGGAACCAGCGCCAAAGTGAACAG 360

Qy   369 GTTGAAGAACAAGATCAACAACACAGGCATCGGAAGAGGCTTTTGTGAAAGAATCCAAG 428
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GTTGAAGAACAAGATCAACAACACAGGGCATCGGAAGAGGCTTTTGTGAAAGAATCCAAG 420

Qy   429 GAGGAGACCCCAGGCACAGAGTGGGAGAAGTGGCCCAGCTGTGTGACTTCAACCCTAAG 488
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GAGGAGACCCCGGGCACAGAGTGGGAGAAGTGGAGCCCAGCTGTGTGACTTCAATCCCAAG 480

Qy   489 AGCAGCAAGCAATGTAAAGACGTGTCCCGGCCTCGGTCGGTCTCAT 535
     |||||||||||||||||||||||||||||||||||||||||||||
Db   481 AGCAGCAAGCAGTGTAAAGATGTGTCCCGCCTGCGCGTCGGTCTCAT 527

RESULT 9
BF467127
LOCUS   BF467127   557 bp   mRNA   linear   EST 04-DEC-2000
DEFINITION   UI-M-CG0p-bre-h-11-0-UI.s1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone
     UI-M-CG0p-bre-h-11-0-UI 3', mRNA sequence.
ACCESSION   BF467127
VERSION   BF467127.1   GI:11536310
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
     Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 557)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE   Normalization and subtraction: two approaches to facilitate gene
     discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
PUBMED   8889548
COMMENT   Contact: Chin, H
     National Institute of Mental Health
     6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
     20892-9643, USA
     Tel: 301 443 1706
     Fax: 301 443 9890
     Email: mBST@mail.nih.gov
     Oligo-dT track not found, Not I site shown in beginning of sequence
     is likely internal to the message. cDNA Library Preparation: M.B.
     Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
     clones from RESEARCH GENETICS. It should be noted that Bento Soares
     is generating a small number of additional specialized
     non-redundant arrays of BMAP cDNAs whose availability will be
     considered under appropriate and limited collaborative arrangements
     Seq primer: M13 Forward
     POLYA=No.
FEATURES   Location/Qualifiers
     source   1..557
       /organism="Mus musculus"
       /mol_type="mRNA"
       /strain="C57BL/6J"

```
                /db_xref="taxon:10090"
                /clone="UI-M-CG0p-bre-h-11-0-UI"
                /clone_lib="DH10B (Life Technologies)"
                /lab_host="NIH_BMAP Ret4 S2"
                /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                polylinker; Site_1: Not I; Site_2: Eco RI; The
                NIH_BMAP Ret4 S2 library is a subtracted library,
                ultimately derived from mouse retina tissue libraries at
                various stages of development. For a detailed description
                of the library from which this clone was derived, please
                visit our web site at brainest.eng.uiowa.edu. The tissue
                for this library was contributed by Dr. Xin-Yuan Fu, Yale
                University School of Medicine
                TAG_SEQ=None found"
```

ORIGIN

```
Query Match           92.2%;  Score 493.2;  DB 2;  Length 557;
Best Local Similarity 96.3%;  Pred. No. 9.2e-111;
Matches 515;  Conservative 0;  Mismatches 19;  Indels 1;  Gaps 1;

Qy    1 CGGCCGCCTTCCTGGCCCAGCAGGAGGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2 CGGCCGCCTTCCTGGCCCAGCAGGAAAGCGAGATTGCTGGCATCGAGAATGACCCGGGTT 61

Qy   61 TCGGGGCACCTGCCGCCCAGCCAGGTGCCCTCTGCGCAGCCCGACTCGCGAGCGGGGGTG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 TCGGGGCACCTGCCGCCCAGCCAGGTGCCCTCTGCGCAGCCCGGACTCGCGAGCGGGGCTG 121

Qy  121 GTTCGGGAGGACATGGGGACTACAGTCAGTGAGATGTGTTCAGGAGGCTAACGGGCCTG 180
          |||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db  122 GTTCAGGAGGACATGAGTGAGTACTACAGTCAGTGAGATGTGTTCAGGAGGCTAACGGGCCTG 181

Qy  181 CCGATGGCTACCGCTGCGATTGCCCAGGCGGACAGGTGACTCAGGAGCCTGAGAGCATCC 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  182 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGCCCAGGAGCCTCAGGAGCATCC 241

Qy  241 GCAAGTGAGAGGAGGAGCAGAGAGAAAGGCTGCCAGGAGTTGATGCGTCGCCTCGAAGGTGA 300
          ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db  242 GGAAGTGGAGAGGAGGAGCAGAGAGAAAGGCTGCCAGGAGTTAGACGCTTCGCCTCGAAGGTGA 301

Qy  301 CCGAACAGGAGTGCGGGGAGAAGCCAAGAAAGACCTCGGAGGAGTGGAACCAGCGCCAAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  302 CCGAACAGGAGTGCGGGGAGAAGCCAAGAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 361

Qy  361 GTGAACAGGTTGAGAAGAACAAGATCAACAACAGGGCATCGGAAGAGGCTTTTGTGAAAG 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  362 GTGAACAGGTTGAGAAGAACAAGATCAACAACAGGGCATCGGAAGAGGCTTTTGTGAAAG 421

Qy  421 AATCCAAGGAGGAGACCCCAGCCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||  |||||||||
Db  422 AATCCAAGGAGGAGACCCCGGCCACAGAGTGGGAGAAGGTGGGAGAAGGT-ACCCAGCTGTGTGACTTCA 480

Qy  481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTGCGTCCGTCGGTGCTCAT 535
          ||||| |||||||||||||||||||||||||| |  ||||||||||||||||||||
Db  481 ATCCCAAGAGCAGCAGCAGTGTAAAGATGTGTCNCGGCTGCGTCCGTCGGTGCTCAT 535
```

RESULT 10

```
BI659300/c
LOCUS       BI659300                 718 bp    mRNA    linear   EST 12-SEP-2001
DEFINITION  603016889F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5347486 5',
            mRNA sequence.
ACCESSION   BI659300
VERSION     BI659300.1  GI:15573536
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 718)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
```

JOURNAL
COMMENT

FEATURES   Location/Qualifiers
     source   1..718
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="NMRI"
                /db_xref="taxon:10090"
                /clone="IMAGE:5347486"
                /tissue_type="tumor, gross tissue"
                /dev_stage="5 months"
                /lab_host="DH10B"
                /clone_lib="NCI_CGAP_Mam4"
                /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
                Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                Library constructed by Life Technologies. Investigators
                providing samples: Lothar Hennighausen/Priscilla Furth,
                NIH Reference for transgenic model: Li et al., Cell Growth
                and Differentiation 7, 3-11 (1996)."

ORIGIN

```
Query Match           91.7%;  Score 490.6;  DB 3;  Length 718;
Best Local Similarity 96.4%;  Pred. No. 4.2e-110;
Matches 502;  Conservative 0;  Mismatches 19;  Indels 0;  Gaps 0;

Qy   15 GCCCAGCAGGAGGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTTCGGGGCACCTGCC 74
          ||||||||||||||||| ||||||||||||||||||||||||  |||||||||||||||
Db  714 GCCCAGCAGGAAAGCGAGATTGCTGGCATCGAGAATGACCCGGGCTTTCGGGGCACCTGCC 655

Qy   75 GCCGCAGGTGCCCTCTGCGCAGCCCGACTCGCGAGCGGGGGTGGTTCGGGGACATG 134
          |||||||||||||||||||||||||||  |||||||||||| |||||||||||||||||
Db  654 GCCGCCAGGTGCCCTCTGCGCAGCCCGGACTCGCGAGCGGGGCTGGTTCAGGAGGACATG 595

Qy  135 GGGACTACAGTCAGTGAGATGTGTTCAGGAGGCTAACGGGCCTGCCGATGGCTACGCT 194
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  594 AGTACTACAGTCAGTGAGATGTGTTCAGGAGGCTAACGGGCCTGCCGATGGCTACGCT 535

Qy  195 GCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCCTCAGGAGCATCCGCAAGTGGAGAG 254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  534 GCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCCTCAGGAGCATCCGGAAGTGGAGAG 475

Qy  255 GAGCAGAGAGAAAGGCTGCCAGGAGTTGATGCGTCGCCTCGAAGGTGACCCAAGAAAGAC 314
          |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db  474 GAGCAGAGAGAAAGGCTGCCAGGAGTTAGACGCTGCCTCGAAGGTGACCCAAGAAAGAC 415

Qy  315 CGGGAGGAAGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAAGTGAACAGGTTGAG 374
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  414 CGGGAGGAAGCCAAGAAAGACCTGGAGGAGTGGAACCAGCGCCAAAGTGAACAGGTTGAG 355

Qy  375 AAGAACAAGATCAACAACAGGGCATCGGAAGAGGCTTTTGTGAAAGAATCCAAGGAGGAG 434
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  354 AAGAACAAGATCAACAACAGGGCATCGGAAGAGGCTTTTGTGAAAGAATCCAAGGAGGAG 295

Qy  435 ACCCCAGGCCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTAAGAGCAGC 494
          |||||| ||||||||||||||||||||||||||||||||||||||||  ||||||||||
Db  294 ACCCCGGCCACAGAGTGGGAGAAGGTAGCCCAGCTGTGTGACTTGACTTCAATCCCAAGAGCAGC 235

Qy  495 AAGCAATGTAAAGACGTGTCCCGCCTGCGTCCGTCGGTGCTCAT 535
          |||| ||||||||||| |  ||||||||||||||||||||||||
Db  234 AAGCAGTGTAAAGATGTGTCGTCCCGGCCTGCGTCCGTCGGTGCTCAT 194
```

```
Db       302  CCGAACAGGAGTGCGGGAGGAAGGCCAAGAAAGACCTGAGGAGTGGAACCAGCGCCAAA  361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       361  GTGAACAGGTTGAGAAGAACAACAGATCAACAACAGGGCATCGAAGAGGCTTTGTGAAAG  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       362  GTGAACAGGTTGAGAAGAACAACAGATCAACAACAGGGCATCGAAGAGGCTNTTGTGAAAG  421

Qy       421  AATCCAAGGAGGAGGACCCCAGGCACAGAGTGGCCCAGCGTGTGTGACTTCA  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       422  AATCCAAGGAGGAGGACCCCGGGCACAGAGTGGGAGAAGGTAGCCCAGCTGTGTGACTTCA  481

Qy       481  ACCTAAGCAGCAAGCAAT  501
              ||||||||||||||||||
Db       482  ATCCCAGAGCAGCAAGCAGT  502
```

```
RESULT 12
BY753851
LOCUS         BY753851                  663 bp    mRNA       linear   EST 17-DEC-2002
DEFINITION    BY753851 RIKEN full-length enriched, osteoclast-like cell Mus
              musculus cDNA clone I420020N05 5', mRNA sequence.
ACCESSION     BY753851
VERSION       BY753851.1  GI:27185506
KEYWORDS      EST.
SOURCE        Mus musculus (house mouse)
  ORGANISM    Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE     1  (bases 1 to 663)
  AUTHORS     Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
              Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
              Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
              Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
              Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
              Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
              Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
              Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
              Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
              Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
              Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
              Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
              Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
              Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
              Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
              Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
              Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
              Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
              Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
              Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
              Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
              Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
              Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
              Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
              Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
              Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
              Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE       Analysis of the mouse transcriptome based on functional annotation
              of 60,770 full-length cDNAs
  JOURNAL     Nature 420, 563-573 (2002)
  PUBMED      12466851
COMMENT       Contact: Yoshihide Hayashizaki
              Laboratory for Genome Exploration Research Group, RIKEN Genomic
              Sciences Center(GSC), Yokohama Institute
              The Institute of Physical and Chemical Research (RIKEN)
              1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
              Tel: 81-45-503-9222
              Fax: 81-45-503-9216
              Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
              Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P.,
              Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F.,
              Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y.,
              Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M.,
```

```
RESULT 11
AI463714
LOCUS         AI463714                  502 bp    mRNA       linear   EST 09-MAR-1999
DEFINITION    AI463714 uc45h03.x1 Soares mammary gland NMLMG Mus musculus cDNA clone
              IMAGE:1400981 3' similar to gb:M20469 CLATHRIN LIGHT CHAIN B
              (HUMAN)., mRNA sequence.
ACCESSION     AI463714
VERSION       AI463714.1  GI:4317744
KEYWORDS      EST.
SOURCE        Mus musculus (house mouse)
  ORGANISM    Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE     1  (bases 1 to 502)
  AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE       Tumor Gene Index
              National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  JOURNAL     Unpublished (1997)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              This clone is available royalty-free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              MGI:912697
              This clone was previously sequenced on the 5' end only, this new
              data is from the 3' end
              Possible reversed clone: polyT not found
              High quality sequence stop: 478.
FEATURES             Location/Qualifiers
     source          1..502
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1400981"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /clone_lib="Soares mammary gland_NMLMG"
                     /note="vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double-stranded  cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector. Library is normalized. Library
                     was constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN

Query Match      88.4%;  Score 472.8;  DB 1;  Length 502;
Best Local Similarity   96.4%;  Pred. No. 9.7e-106;
Matches  483;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps      0;

Qy       1  CGGGCGCCTTCCTGCCCAGCAGGAGGAGCCAGATTGCTCGGCATCGAGAATGACTCGGGTT   60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2  CGGCGCCGCCTTCCTGCCCAGCAGGAGGAGCCAGATTGCTCGGCATCGAGAATGACCCGGGTT   61

Qy      61  TCGGGGCACCTGCCGCCAGCCAGGTGGCCTTCTGCGCAGCCCGGACTCGGGACTCGGGGGTG  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62  TCGGGGCACCTGCCGCCAGCCAGGTGGCCTTCTGCGCCCAGCCCGGACTCGGGACTCGGGGGTG  121

Qy     121  GTTCGGAGGACATGGGAGCTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGCCTG  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122  GTTCAGGAGGACATGGGAGCTACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGCCTG  181

Qy     181  CCGATGGCTACGCTGCGATTGCCCAGGCCGACAGGTTGACTCAGGAGCCTGAGAGCATCC  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182  CCGATGGCTACGCTGCGATTGCCCAGGCCGACAGGTTGACTCAGGAGCCTGAGAGCATCC  241

Qy     241  GCAAGTGGAGAGAGCAGAAGAAAGGCTGCAGGAGTTGGATGCTCGCTCGAAGGTGA  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242  GGAAGTGGAGAGAGCAGAAGAAAGGCTGCAGGAGTTAGACCTGCCTCGAAGGTGA  301

Qy     301  CCGAACAGCAGTGGCGGGGAGGAGAAGGCCAAAAAAAGACCTGAGGAGTGGAACCAGCGCCAAA  360
```

```
Qy   481  ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCGCCTCGGTGCTCAT  535
               ||||||| ||||||| ||||||||||||||||||||||||||||
Db   608  ATCC-CAGAGCAGNCAGCAGTGTAAAGATGTGTCCGCCTCGGTGCTCAT  661
```

RESULT 13
BB854705/c

LOCUS   BB854705   569 bp   mRNA   linear   EST 27-SEP-2000
DEFINITION   ux32h10.y1 Soares_NMMAX_maxillary_process Mus musculus cDNA clone
             IMAGE:3512035 5' similar to SW:CLCB_RAT P08082 CLATHRIN LIGHT CHAIN
             B.; mRNA sequence.
ACCESSION   BB854705
VERSION   BB854705.1 GI:10313317
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 569)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT   Other ESTs: ux32h10.x1
             Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             This clone is available royalty-free through LLNL; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:1394915
             Possible reversed clone: similarity on wrong strand
             Seq primer: -40RP from Gibco
             High quality sequence stop: 300.
FEATURES   Location/Qualifiers
     source   1..569
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3512035"
                     /tissue_type="maxillary process"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="Soares_NMMAX_maxillary_process"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site 1: NotI; Site 2: EcoRI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCGTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN
Query Match   88.1%; Score 471.4; DB 2; Length 569;
Best Local Similarity 93.2%; Pred. No. 2.2e-105;
Matches   493; Conservative   0; Mismatches 36; Indels   0; Gaps   0;

```
Qy    7  CCTTCCTGCGCCAGCAGGAGGAGGCGAGATTGCTGCGCATCGAGAATGACTCGGGTTCGGGG   66
              ||||||| | |||||| |||||||| |||||||||||| |||||||| || |||||| ||||
Db   565  CCTTCCTGTCCTAGCAGGAAAGCGAGAGTTTGCTGCGCATGGAGAATGATGACCTGGTTTTTTGGGG  506

Qy   67  CACCTGCGCCAGCCAGGTGGCCTCTGCGCCAGCCAGCCGGTACTCGCCGGACGAGGCGTGTTCGG  126
              |||| |||| ||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   505  CACCAGCGCCCAGCCAGCCAGGTGGCCTCTGCGCCAGCCAGCCGGTACTCGCCGGACGAGGCGTGTTCAT  446

Qy  127  AGGACATGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTTTTGTGAAAG  186
              || ||||| | |||||||||||||| ||||||||||||||||||||||| ||
Db   445  AGTACATGAGTACTACAGTCAATGAAGATGTGTTTCAGGAGGCTTTTGTGANAG  386

Qy  187  GCTACGCTGCGATTGCCCAGGTCCGGACACAGTTGACTCAGTGTGTGACAAGT  246
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   385  GCTACGCTGCGATTGCCCAGTTCGGACACAGTTGACTCAGTGTGTGACTTCA  326
```

Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N.,
Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M.,
Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
  Computational Analysis of Full-Length Mouse cDNAs Compared with
  Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
  Normalization and subtraction of cap-trapper-selected cDNAs to
  prepare full-length cDNA libraries for rapid discovery of new
  genes. Genome Res. 10 (10), 1617-1630 (2000)
  RIKEN integrated sequence analysis (RISA) system--384-format
  sequencing pipeline with 384 multicapillary sequencer. Genome Res.
  10 (11), 1757-1771 (2000)
  Computer-based methods for the mouse full-length cDNA
  encyclopedia: real-time sequence clustering for construction of a
  nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
  cDNA library was prepared and sequenced in Mouse Genome
  Encyclopedia Project of Genome Exploration Research Group in Riken
  Genomic Sciences Center and Genome Science Laboratory in RIKEN.
  Division of Experimental Animal Research in Riken contributed to
  prepare mouse tissues.
  Tissues were provided by Takashi Ishikawa (Department of Surgery
  2 Yokohama City University 3-9 Fukuura,Kanazawa-ku,Yokohama
  236-0004 Japan) whose assistance we gratefully acknowledge.
  Please visit our web site (http://genome.gsc.riken.go.jp) for
  further details.
FEATURES   Location/Qualifiers
     source   1..663
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="1420020N05"
                     /clone_type="osteoclast-like cell"
                     /cell_type="osteoclast-like cell"
                     /clone_lib="RIKEN full-length enriched, osteoclast-like
                     cell"

ORIGIN
Query Match   88.2%; Score 471.8; DB 5; Length 663;
Best Local Similarity 94.8%; Pred. No. 1.8e-105;
Matches   507; Conservative   0; Mismatches 26; Indels   2; Gaps   2;

```
Qy    1  CGGCCGGCTTCCTGGCCCAGCAGGAGGAGGCGAGATTGCTGGCATCGAGAATGACTCGGGTT   60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   129  CGGCCGGCTTCCTGGCCCAGCAGGAGGAGGCGAGATTGCTGGCATCGAGAATGACCCGGGTT  188

Qy   61  TCGGGGCACCTGCGCCGCCCAGCCAGGTGGCCTCTGCGCAGCCCGACTCGCGAGCGGGGGTG  120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   189  TCGGGGCACCTGCGCCGCCCAGCCAGGTGGCCTCTGCGCAGCCCGACTCGCGAGCGGGGCTG  248

Qy  121  GTTCGGAGGACATGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACGGGCCTG  180
              |||| ||||||||| | |||||||||||||||||||||||||||||||||||||||||
Db   249  GTTCAGAGGACATGAGTACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACGGGCCTG  308

Qy  181  CCGATGGCTACGCTGCGATTGCCCAGGAGGCGGGACAGGTTGACTCAGGAGAGCATCC  240
              ||||||||||||||||||||||||||||| |||||| |||||||||||||||||||||
Db   309  CCGATGGCTACGCTGCGATTGCCCAGGCGGCAGGTTGACTCAGGAGAGCATCC  368

Qy  241  GCAAGTGAGAGAGGAGGAGAGAAAAGGCTCGCAGGAGGTTGGATGCTCGCTCGAAGGTGA  300
              | |||||||||||||||||||||||| || ||||||| || |||| |||| |||||||||
Db   369  GGAAGTGAGAGAGGAGGAGAGAAAGAAAGGCTTAGACGCTGCCTCGCAGGAGGTGA  428

Qy  301  CCGAACACAGAGAGTGTGCGGGGAGAAGGCCAAAAAAAGACCTGGAGGAGTGAACCAGCGCCAAA  360
              |||||||||||||||| |||||||||||||||| |||||||||||||||||||||||| |||||
Db   429  CCGAACACAGAGAGTGCGGGGAGAAGGCCAAGAAAGACCTGGAGGAGTGAACCAGCCCAAA  488

Qy  361  GTGAACACAGGTTGAGAAGAAGACAAGATCAACAACAGGGCATCGGAAGAGGCTTTTGTGAAAG  420
              ||||||||||||||| ||||| | | |||||| |||||||||||| |||||||||||||| ||
Db   489  GTGAACACAGGTTGAAGAAGAACTAGATCAANCACAGGGCATCGGAAGAGGCTTTTGTGANAG  548

Qy  421  AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCA  480
              || |||| ||||||||| |||| ||||||||||||||||||| ||||||||||||||||||
Db   549  AANTCCAGGAGGAGACACCCC-GGCACAGAGTGGGAGAAGGTAGCCCAGCTGTGTGACTTCA  607
```

Best Local Similarity   96.6%;   Pred. No. 2.1e-104;
Matches 477; Conservative   0; Mismatches 17; Indels   0; Gaps   0;

```
Qy     1  CGGCCGCCTTCCTGCCCCAGCAGGAGAGCGGAGTTGGATGCTGCCTCGAAGGTGACCGAAC   60
          ||||||||||| |||||||||||||||||||  ||||||||||||||||| ||||||||||
Db     3  CGGCCGCCTTCCTGCCCCAGCAGGAGAGCGGAGTTGGATGCTGCCTCGAAGGTGACCGGTT   62

Qy    61  TCGGGGCACCTGCCCAGCCAGGTGGCCTCTCGCGAGACTCGGCCGGAGCGCGGGGGTG   120
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    63  TCGGGGCACCTGCCCAGCCAGGTGGCCTCTCGCGAGACTCGGCCGGAGCGCGGGGGCTG   122

Qy   121  GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG   180
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   123  GTTCAGAGGACATGAGTACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG   182

Qy   181  CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCCTCAGGAGCATCC   240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   183  CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCCCCAGGAGCATCC   242

Qy   241  GCAAGTGGAGAGAGCAGAGAAGGCTGCAGAAGAAAAGGCTGCAGGAGTGGATGCTGCCTCGAAGGTGA   300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   243  GGAAGTGGAGAGAGAGAGCAGAGAAGGCAGAAGAAAAGGCTGCAGGAGTTAGACGCTGCCTCGAAGGTGA   302

Qy   301  CCGAACAGGAGTGCGGGGAAGAGACCTGGAGGAGTGGAACCAGCCGCCAAA   360
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   303  CCGAACAGGAGTGCGGGGGAAGAGGCCAAGAAAGACCTGGAGGAGTGGAACCAGCCGCCAAA   362

Qy   361  GTGAACAGGTTGAGAAGAACAAGATCAACAACACAGGGCATCGGAAGAGGCTTTGTGAAAG   420
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   363  GTGAACAGGTTGAGAAGAACAAGATCAACAACACAGGGCATCGGAAGAGGCTNTTGTGAAAG   422

Qy   421  AATCCAAGAGGAGACCCCAGGCACAGAGTGTGGGAAGAGGGTGCCCAGCTGTGTGACTTCA   480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   423  AATCCAAGAGGAGACCCCGGGCACAGAGTGTGGGAGAGGTAGCCCAGCTGTGTGACTTCA   482

Qy   481  ACCCTAAGAGCAGC   494
          |||||| |||||||
Db   483  ATCCCAAGAGCAGC   496
```

RESULT 15
AI104105
LOCUS   AI104105   476 bp   mRNA   linear   EST 31-JAN-1999
DEFINITION EST213394 Normalized rat heart, Bento Soares Rattus sp. cDNA clone
            RHEBX34 3' end, mRNA sequence.
ACCESSION   AI104105
VERSION   AI104105.1 GI:3708518
KEYWORDS   EST.
SOURCE   Rattus sp.
  ORGANISM Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 476)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE   Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT   Other_ESTs: TC52991
            Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES   Location/Qualifiers
     source   1..476
                     /organism="Rattus sp."
                     /mol_type="mRNA"
                     /db_xref="ATCC (inhost):2025352"
                     /db_xref="taxon:10118"

```
Qy   247  GGAGAGGAGGAGCAGAAGAAAGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGACCGAAC   306
          |||||||||||||||||||||||| || |||||||||||||||||||||||||||||||
Db   325  GGAGAGGAGGAGCAGAAGAAAGCTGCAGGAGTTAGACGCTGCCTCGAAGGTGACCGAAC   266

Qy   307  AGGAGTGCGGGGAAGAAGGCCAAAAAAGACCTGAGGAGTGGAACCAGCCGCCAAAGTGAAC   366
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   265  AGGAGTGCGGGGGAAGAAGACCTGGAGGAGTGGAACCAGCCGCCAAAGTTAAC   206

Qy   367  AGGTTGAGAAGAACAACAACAGGGCATCGGAAGAGGCTTTTGTGAAAGAATCCA   426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   205  AGGTTGAGAAGAACAACAACAGGGCATCGGAAGAGGGCTTTTGTGAAAGAATCCA   146

Qy   427  AGGAGGAGCCCCAGGCACAGAGTGGGGAGAAGGTGGCCCAGCTGTGTGACTTCAAACCCTA   486
          ||||||||||||||||||||||||||||||||||||||||||||||| |||||||| ||
Db   145  AGGAGGAGCCCCCGGGCCACAGAGTGTGGGAGAAGGTAGCCCAGCTGTGTGACTTCAATCCCA   86

Qy   487  AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTCGGTCGGTGCTCAT   535
          ||||||||||||| |||| |||||||||||||||||||||||||||||
Db    85  AGAGCAGCAAGCAGTGTAAAGATGTGTCCCGCCTCGGTCGGTGCTCAT   37
```

RESULT 14
BF227172
LOCUS   BF227172   496 bp   mRNA   linear   EST 29-DEC-2000
DEFINITION uz20f03.x1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3669629 3'
            similar to SW:CLCB_RAT P08082 CLATHRIN LIGHT CHAIN B ;, mRNA
            sequence.
ACCESSION   BF227172
VERSION   BF227172.1 GI:11135084
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 496)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT   Other_ESTs: uz20f03.y1
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml
            MGI:1430397
            Possible reversed clone: similarity on wrong strand
            High quality sequence stop: 448.
FEATURES   Location/Qualifiers
     source   1..496
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="mix FVB/N, C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3669629"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="7 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam5"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Robin Humphreys,
                     NIH"

ORIGIN

Query Match   87.4%;   Score 467.4;   DB 2;   Length 496;

```
              /clone="RHEBX34"
              /clone_lib="Normalized rat heart, Bento Soares"
              /note="Organ: heart; Vector: pT7T3Pac; Site_1: EcoRI;
              Site_2: NotI"

ORIGIN

Query Match           87.3%;  Score 467.2;  DB 1;  Length 476;
Best Local Similarity  99.4%;  Pred. No. 2.3e-104;
Matches  469;  Conservative  0;  Mismatches  0;  Indels  3;  Gaps  0;

Qy    46 AGAATGACTCGGGTTTCGGGGCACCTGCCGCCAGCCAGGTGGCCTCTGCGCCAGCCCCGAC 105
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 AGAATGACTCGGGTTTCGGGGCACCTGCCGCCAGCCAGGTGGCCTCTGCGCCAGCCCCGAC  60

Qy   106 TCGCGAGCGGGGGTGGTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGG 165
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TCGCGAGCGGGGGTGGTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGG 120

Qy   166 AGGCTAACGGGCCTGCCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGG 225
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGGCTAACGGGCCTGCCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGG 180

Qy   226 AGCCTGAGAGCATCCGCAAGTGAGAGAGGAGGAGAGAAAAGGCTGCAGGAGTTGGATG 285
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AGCCTGAGAGCATCCGCAAGTGAGAGAGGAGGAGAGAAAAGGCTGCAGGAGTTGGATG 240

Qy   286 CTGCCTCGAAGTGACCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGT 345
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CTGCCTCGAAGTGACCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGT 300

Qy   346 GGAACCAGCGCGCCAAAGTGAACAGGTTGAGAAGAAACAAGATCAACAACAGGGCATCCGGAAG 405
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GGAACCAGCGCGCCAAAGTGAACAGGTTGAGAAGAAACAAGATCAACAACAGGGCATCCGGAAG 360

Qy   406 AGGCTTTTGTGAAAGAATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCC 465
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AGGCTTTTGTGAAAGAATCCAAGGAGGAGACCCCAGGCACAGAGTGGTAGAAGGTGGCCC 420

Qy   466 AGCTGTGTGACTTCAACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCG 517
         |||||||||||||||||||||||||||||||||||||||||||| || ||
Db   421 AGCTGTGTGACTTCAACCCTAAGAGCAGCAAGCAATGTAAAGACGTGCCTCG 472
```

Search completed: January 5, 2006, 21:32:34
Job time : 3756 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    January 5, 2006, 19:27:35  ; Search time 159 Seconds
                                       (without alignments)
                                       5981.104 Million cell updates/sec

Title:           US-10-664-705-145
Perfect score:   535
Sequence:        1 cggccgccttcctggcccag........cgcctgcgctcggtgctcat 535

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters:     2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    Issued Patents NA:*
       1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
       2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
       3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
       4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
       5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
       6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
       7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
       8: /cgn2_6/ptodata/1/ina/RB_COMB.seq:*
       9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
c	1	78	14.6	7218	2	US-08-232-463-14	Sequence 14, Appl
c	2	65.2	12.2	658	3	US-09-385-982-327	Sequence 327, App
c	3	55.8	10.4	12695	3	US-09-949-016-16775	Sequence 16775, A
c	4	54.2	10.1	51259	3	US-08-781-891-209	Sequence 209, App
c	5	54.2	10.1	51259	3	US-09-618-166-209	Sequence 209, App
c	6	52.8	9.9	16442	3	US-08-781-891-208	Sequence 208, App
c	7	52.8	9.9	16442	3	US-09-618-166-208	Sequence 208, App
c	8	50.6	9.5	1926	3	US-09-249-585A-2	Sequence 2, Appli
c	9	50.6	9.5	1926	3	US-09-410-399-3	Sequence 3, Appli
	10	50.6	9.5	2580	3	US-09-050-863-2	Sequence 2, Appli
	11	50.6	9.5	2580	3	US-09-359-081-2	Sequence 2, Appli
	12	50.6	9.5	5452	2	US-09-130-114-1	Sequence 1, Appli
c	13	50.6	9.5	8705	3	US-09-647-344A-14	Sequence 14, Appli
c	14	50.6	9.5	9600	3	US-08-910-647-1	Sequence 1, Appli
	15	50.6	9.5	9600	3	US-09-620-925-1	Sequence 1, Appli
	16	50.6	9.5	10596	2	US-07-984-811-15	Sequence 15, Appl
	17	50.6	9.5	10596	2	US-07-885-971-15	Sequence 15, Appl
	18	50.6	9.5	10596	2	US-08-087-783A-15	Sequence 15, Appl
	19	50.6	9.5	10596	2	US-08-194-088B-15	Sequence 15, Appl
	20	50.6	9.5	10596	2	US-08-194-087-15	Sequence 15, Appl
	21	50.6	9.5	10596	6	PCT-US93-04648-15	Sequence 15, Appl
c	22	50.6	9.5	16080	3	US-09-724-566A-48	Sequence 48, Appl
c	23	50.6	9.5	16080	3	US-09-471-669A-48	Sequence 48, Appl
c	24	48.6	9.1	150394	3	US-09-949-016-13042	Sequence 13042, A
	25	48.4	9.0	3489	2	US-08-728-323A-1	Sequence 1, Appli
	26	48.4	9.0	3489	3	US-09-298-568-1	Sequence 1, Appli
	27	48.4	9.0	3489	3	US-09-410-399-1	Sequence 1, Appli
	28	48.4	9.0	3489	2	US-09-894-273-1	Sequence 1, Appli
c	29	48.4	9.0	32207	3	US-08-770-379-20	Sequence 20, Appl
c	30	48.4	9.0	32207	3	US-08-757-669A-20	Sequence 20, Appl
c	31	48.4	9.0	32207	3	US-09-230-371A-20	Sequence 20, Appl
	32	47.8	8.9	669	3	US-09-669-751-79	Sequence 79, Appl
c	33	47.8	8.9	254964	3	US-09-949-016-12583	Sequence 12583, A
c	34	47.8	8.9	254964	3	US-09-949-016-17392	Sequence 17392, A
	35	47.4	8.9	2093	3	US-10-104-047-1666	Sequence 1666, Ap
	36	47.4	8.9	767677	3	US-09-949-016-12147	Sequence 12147, A
	37	47.4	8.9	767677	3	US-09-949-016-17361	Sequence 17361, A
c	38	47.2	8.8	85850	3	US-09-949-016-13424	Sequence 13424, A
	39	46.8	8.7	4092	3	US-09-566-921-115	Sequence 115, App
c	40	46.2	8.6	72549	3	US-09-949-016-16477	Sequence 16477, A
c	41	44.8	8.4	1926	2	US-09-249-585A-4	Sequence 4, Appli
c	42	44.8	8.4	1931	2	US-09-130-114-2	Sequence 2, Appli
	43	44.6	8.3	289	3	US-09-007-005-17	Sequence 17, Appl
	44	44.6	8.3	289	3	US-09-244-796-17	Sequence 17, Appl
	45	44	8.2	4722	3	US-08-979-608A-14	Sequence 14, Appl

ALIGNMENTS

```
RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
;   APPLICANT: DORNER, F.
;   APPLICANT: SCHEIFLINGER, F.
;   APPLICANT: FALKNER, F. G.
;   TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
;   NUMBER OF SEQUENCES: 52
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Foley & Lardner
;     STREET: 1800 Diagonal Road, Suite 500
;     CITY: Alexandria
;     STATE: VA
;     COUNTRY: USA
;     ZIP: 22313-0299
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/232,463
;     FILING DATE:
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/07/935,313
;     FILING DATE:
;     APPLICATION NUMBER: EP 91 114 300.6
;     FILING DATE: 26-AUG-1991
;   ATTORNEY/AGENT INFORMATION:
;     NAME: BENT, Stephen A.
;     REGISTRATION NUMBER: 29,768
;     REFERENCE/DOCKET NUMBER: 30472/114 IMMU
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (703)836-9300
;     TELEFAX: (703)683-4109
;     TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 7218 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;     CLONE: pTZgpt-F1s
```

US-08-232-463-14

```
Query Match      14.6%;  Score 78;  DB 2;  Length 7218;
Best Local Similarity   5.7%;  Pred. No. 3.9e-11;
Matches   24;  Conservative  245;  Mismatches  155;  Indels   0;  Gaps   0;

Qy      78 AGCCAGGTGGCCTCTGCGGACTCGCGAGCCGGACTCGGCGAGCGGGGGTGGTTCGGAGGACATGGGG 137
           ||||      |||  :: ::  :::  :::  :::::: ::: :::::
Db    1453 AGATAGAAGAAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1394

Qy     138 ACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTGCCGATGGCTACGCTGCG 197
           : : ::: :::::: :  : :::::: :: :::::: :: : : ::::
Db    1393 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1334

Qy     198 ATTGCCCAGGCCGGACCAGGTTGACTCAGGAGCCTGAGAGCATCCGCAGTGGAGAGAGGAG 257
           : : :::    :  :: :::  :  ::::: : ::::::
Db    1333 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1274

Qy     258 CAGAAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAGGTGACCGAACAGGAGTGGCGG 317
           :::::::::::: ::: :: :::: :: ::::::  ::::: :: :::::: :: ::: ::
Db    1273 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1214

Qy     318 GAGAAGGCCAAAAAAGACCTGGAGAGTGGAACCAGCGCCAAAGTGAACAGGTTGAGAAG 377
           :::::::: :  ::::::::  :::: :: ::::::::: ::::::::::::::::::::
Db    1213 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1154

Qy     378 AACAAGATCAACAACAGCGGCATCGGAAGAGGCTTTTGTGAAAGAATCCAAGGAGGACC 437
           : :: ::::::::::::::::::::::: :::::
Db    1153 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1094

Qy     438 CCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTAAGAGCAGCAAG 497
           ::: : ::::::: :::::: ::  :::
Db    1093 RRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCTCCCTGACCTGCAGCCAAGCTCG 1034

Qy     498 CAAT 501
           |||
Db    1033 GAAT 1030
```

RESULT 2
US-09-385-982-327/c
; Sequence 327, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
;   APPLICANT: ENDEGE, WILSON O., ET AL.
;   TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;   TITLE OF INVENTION: PRODUCTS II
;   FILE REFERENCE: CCDNA-260XX
;   CURRENT APPLICATION NUMBER: US/09/385,982
;   CURRENT FILING DATE: 1999-08-30
;   EARLIER APPLICATION NUMBER: 09/328,111
;   EARLIER FILING DATE: 1999-06-08
;   EARLIER APPLICATION NUMBER: 60/117,393
;   EARLIER FILING DATE: 1999-01-27
;   EARLIER APPLICATION NUMBER: 60/098,639
;   EARLIER FILING DATE: 1998-08-31
;   NUMBER OF SEQ ID NOS: 544
;   SOFTWARE: FastSEQ for Windows Version 3.0
;   SEQ ID NO 327
;   LENGTH: 658
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(658)
;   OTHER INFORMATION: n = A,T,C or G
US-09-385-982-327

```
Query Match      12.2%;  Score 65.2;  DB 3;  Length 658;
Best Local Similarity   56.2%;  Pred. No. 5.1e-08;
Matches  114;  Conservative   0;  Mismatches  88;  Indels   1;  Gaps   1;

Qy     321 AAGGCCAAAAAAGACCTGGAGAGTGGAACCAGCGCCAAAGTGAACAGGTTGAGAAGAAC 380
```

; Sequence 16775, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 16775
;   LENGTH: 12695;
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-16775

RESULT 3
US-09-949-016-16775/c

```
Db     564 AAGGCCCATTAAGCNGTTANAAANAATGGTTNCAAGNCAGGNCGACNCAGTTCCAGAA-ACC 506
Qy     381 AAGATCAACAACAGGGCATCGGAAGAGGCTTTTGTGAAAGAATCCAAGGAGGACCCCA 440
Db     505 AAAANCAACCACCAAGGCANCACAAAAAAANCNTTNGGAAANGNCATTGACAANTGGTCCCCA 446
Qy     441 GGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTAAGAGCAGCAAGCAA 500
Db     445 GGCNNTNAGTGGAACGGCGGTGGCCCGGNNNGTGTGCCTTTANCCCCAAGTTTAGCAAGCAG 386
Qy     501 TGTAAAGACGTGTCCCGCCTGCG 523
Db     385 GCCAAAGATNTTTCCCGNATGNG 363
```

```
Query Match      10.4%;  Score 55.8;  DB 3;  Length 12695;
Best Local Similarity   53.4%;  Pred. No. 5e-05;
Matches  117;  Conservative   0;  Mismatches  102;  Indels   0;  Gaps   0;

Qy     244 AGTGGAGAGGAGGCAGAGAAAAGGCTGCAGAGTTGGATGCTGCCTCGAAGGTGACCG 303
Db    9676 AGGAGGAACGAGGAGGAGGAGGAGAGGAGGAGGAGCAGGAGGAAGAAGGAGGAGGAGG 9617
Qy     304 AACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGAGTGAACCAGCGCCAAAGTG 363
Db    9616 AAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9557
Qy     364 AACAGGTTGAGAAGAACAACAGATCAACAACAGGGCATCGGAGGAGGAGGAGGAGGAAGAAT 423
Db    9556 AAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9497
Qy     424 CCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGG 462
Db    9496 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9458
```

RESULT 4
US-08-781-891-209/c
; Sequence 209, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
;   APPLICANT: Fu, Ying-Hui
;   APPLICANT: Yu, Chang-En
;   APPLICANT: Oshima, Junko
;   APPLICANT: Mulligan, John T.
;   APPLICANT: Schellenberg, Gerald D.
;   TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO

TITLE OF INVENTION: WERNER'S SYNDROME
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,891
FILING DATE: 27-DEC-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: No. 6090620tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 240052.419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 209:
SEQUENCE CHARACTERISTICS:
LENGTH: 51259 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-781-891-209:

Query Match   10.1%;   Score 54.2;   DB 3;   Length 51259;
Best Local Similarity 51.9%;   Pred. No. 0.00022;
Matches 122; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

Qy   236 CATCCGCAAGTGGAGAGGAGGAGGAGCAGAAGAAAAGGCTCGCAGGAGTTGGATGCTGCCTCGAA 295
              ||| ||||| |||||||||||||||||| |||||||
Db   269 CATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAGGA 210

Qy   296 GGTGACCGAACAGGAGTGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTTGGAACCAGCG 355
Db   209 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAGGA 150

Qy   356 CCAAAGTGAACAGGTTGAGAGAACAAGATCAACACAGGGCCATCGGAGGAGGCTTTTGT 415
Db   149 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 90

Qy   416 GAAAGAATCCAAGGAGGAGGAGGAGCAACAGATCAACACAGGGCCAGGCACAGAGTGGGAGAAGTGGCCCAGCTG 470
Db   89 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGGCCAGGAGGAGGAGGAGCAG 35

RESULT 5
US-09-618-166-209/c
; Sequence 209, Application US/09618166
; Patent No. 6583112
; GENERAL INFORMATION:
APPLICANT: Fu, Ying-Hui
           Yu, Chang-En
           Oshima, Junko
           Mulligan, John T.
           Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
                    WERNER'S SYNDROME
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed Intellectual Property Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA

ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/618,166
FILING DATE: 17-Jul-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: McMasters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 240052.419C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 209:
SEQUENCE CHARACTERISTICS:
LENGTH: 51259 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 209:
US-09-618-166-209

Query Match   10.1%;   Score 54.2;   DB 3;   Length 51259;
Best Local Similarity 51.9%;   Pred. No. 0.00022;
Matches 122; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

Qy   236 CATCCGCAAGTGGAGAGGAGGAGGAGCAGAAGAAAAGGCTCGCAGGAGTTGGATGCTGCCTCGAA 295
              ||| ||||| |||||||||||||||||| |||||||
Db   269 CATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAGGA 210

Qy   296 GGTGACCGAACAGGAGTGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTTGGAACCAGCG 355
Db   209 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAGGA 150

Qy   356 CCAAAGTGAACAGGTTGAGAGAACAAGATCAACACAGGGCCATCGGAGGAGGCTTTTGT 415
Db   149 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 90

Qy   416 GAAAGAATCCAAGGAGGAGGAGGAGCAACAGATCAACACAGGGCCAGGCACAGAGTGGGAGAAGTGGCCCAGCTG 470
Db   89 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGGCCAGGAGGAGGAGGAGCAG 35

RESULT 6
US-08-781-891-208/c
; Sequence 208, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
APPLICANT: Fu, Ying-Hui
APPLICANT: Yu, Chang-En
APPLICANT: Oshima, Junko
APPLICANT: Mulligan, John T.
APPLICANT: Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
TITLE OF INVENTION: WERNER'S SYNDROME
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```
;
;
;
;
;      APPLICATION NUMBER: US/08/781,891
;      FILING DATE: 27-DEC-1996
;      CLASSIFICATION: 800
;      ATTORNEY/AGENT INFORMATION:
;      NAME: No. 6090620tenburg Ph.D., Carol
;      REGISTRATION NUMBER: 39,317
;      REFERENCE/DOCKET NUMBER: 240052.419
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (206) 622-4900
;      TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 208:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 16442 base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
; US-08-781-891-208

Query Match      9.9%;    Score 52.8;  DB 3;  Length 16442;
Best Local Similarity    48.4%;  Pred. NO. 0.00036;
Matches 147;  Conservative  0;  Mismatches 157;  Indels  0;  Gaps  0;

Qy      205  AGGCGGACAGGTTGACTTCAGGAGCCTGAGAGCATCCGCAAGTGGAGGAGGAGCAGAAGA  264
             ||| ||    ||| ||||    |||    ||||   ||||||||||||||||||
Db     16410 AGGAGGAGGAGCAGGAGGAGCAGGAGCAGGAGGAGGAGCAGGAGGAGGAGGAGGAGGAGC  16351

Qy      265  AAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGTGACCGAACACAGGAGTGGCGGGAGAAGG  324
             | ||||    ||||    |||||    || ||||  |||||||||||||||||
Db     16350 AGGAGGCAGGAGGAGCAGGAGGAGCAGGAGGAGCAGGAGGAGGAGCAGGAGGAGGAGGAGC  16291

Qy      325  CCAAAAAAGACCTGGAGGAGTGAACCAGCGCCAAAGTGAACAGGTTGAGAAGAACAAGA  384
             |     | || ||||   || ||||||    ||||||||||||||||||||||
Db     16290 AGGAGGAGGAGCAGGAGGAGCAGGAGGAGCAGGAGGAGGAGCAGGAGGAGGAGGAGCAGG  16231

Qy      385  TCAACAACAGGGCATCCGAAGAGAGCGTTTTGTGAAAGAATCCAAGGAGGAGACCCCAGGCA  444
             |||    |||||    ||| |  ||    ||||||||||||
Db     16230 AGCAGGAGGAGGAGGAGGAGGAGTAGAAGAAGAGGAGGAAGAAGAAGAAGAAGAAGAAGA  16171

Qy      445  CAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTAAGAGCAGCAAGCAATGTA  504
             ||||| ||| || ||     ||| |||  |||  || |||
Db     16170 AGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA  16111

Qy      505  AAGA  508
             || |
Db     16110 AATA  16107

RESULT 7
US-09-618-166-208/c
; Sequence 208, Application US/09618166
; Patent No. 6583112
; GENERAL INFORMATION:
;      APPLICANT: Fu, Ying-Hui
;               Yu, Chang-En
;               Oshima, Junko
;               Mulligan, John T.
;               Schellenberg, Gerald D.
;      TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
;               WERNER'S SYNDROME
;      NUMBER OF SEQUENCES: 209
;      CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Seed Intellectual Property Law Group
;      STREET: 701 Fifth Avenue, Suite 6300
;      CITY: Seattle
;      STATE: Washington
;      COUNTRY: USA
;      ZIP: 98104-7092
;      COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.30
;      CURRENT APPLICATION DATA:
```

```
;
;
;
;
;      APPLICATION NUMBER: US/09/618,166
;      FILING DATE: 17-Jul-2000
;      CLASSIFICATION: <Unknown>
;      ATTORNEY/AGENT INFORMATION:
;      NAME: McMasters, David D,
;      REGISTRATION NUMBER: 33,963
;      REFERENCE/DOCKET NUMBER: 240052.419C1
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (206) 622-4900
;      TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 208:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 16442 base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;      SEQUENCE DESCRIPTION: SEQ ID NO: 208:
; US-09-618-166-208

Query Match      9.9%;    Score 52.8;  DB 3;  Length 16442;
Best Local Similarity    48.4%;  Pred. No. 0.00036;
Matches 147;  Conservative  0;  Mismatches 157;  Indels  0;  Gaps  0;

Qy      205  AGGCGGACAGGTTGACTTCAGGAGCCTGAGAGCATCCGCAAGTGGAGGAGGAGCAGAAGA  264
             ||| ||    ||| ||||    |||    ||||   ||||||||||||||||||
Db     16410 AGGAGGAGGAGCAGGAGGAGCAGGAGCAGGAGGAGGAGCAGGAGGAGGAGGAGGAGGAGC  16351

Qy      265  AAAGGCTGCAGGAGTTGGATGCTGTCGCCTCGAAGTGACCGAACACAGGAGTGGCGGGAGAAGG  324
             | ||||    ||||    |||||    || ||||  |||||||||||||||||
Db     16350 AGGAGGCAGGAGGAGCAGGAGGAGCAGGAGGAGCAGGAGGAGGAGCAGGAGGAGGAGGAGC  16291

Qy      325  CCAAAAAAGACCTGGAGGAGTGAACCAGCGCCAAAGTGAACAGGTTGAGAAGAACAAGA  384
             |     | || ||||   || ||||||    ||||||||||||||||||||||
Db     16290 AGGAGGAGGAGCAGGAGGAGCAGGAGGAGCAGGAGGAGGAGCAGGAGGAGGAGGAGCAGG  16231

Qy      385  TCAACAACAGGGCATCCGAAGAGAGGCTTTTGTGAAAGAATCCAAGGAGGAGACCCCAGGCA  444
             |||    |||||    ||| |  ||    ||||||||||||
Db     16230 AGCAGGAGGAGGAGGAGGAGGAGTAGAAGAAGAGGAGGAAGAAGAAGAAGAAGAAGAAGA  16171

Qy      445  CAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTAAGAGCAGCAAGCAATGTA  504
             ||||| ||| || ||     ||| |||  |||  || |||
Db     16170 AGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA  16111

Qy      505  AAGA  508
             || |
Db     16110 AATA  16107

RESULT 8
US-09-249-585A-2
; Sequence 2, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
;      APPLICANT: Horlick, Robert
;      TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
;      FILE REFERENCE: 0867/0D905
;      CURRENT APPLICATION NUMBER: US/09/249,585A
;      CURRENT FILING DATE: 1999-02-11
;      NUMBER OF SEQ ID NOS: 18
;      SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;      LENGTH: 1926
;      TYPE: DNA
;      ORGANISM: Epstein Barr Virus
;      FEATURE:
;      NAME/KEY: CDS
;      LOCATION: (1)..(1926)
;      OTHER INFORMATION: coding strand of EBNA-1 DNA
; US-09-249-585A-2

Query Match      9.5%;    Score 50.6;  DB 3;  Length 1926;
Best Local Similarity    46.9%;  Pred. No. 0.00068;
Matches 158;  Conservative  0;  Mismatches 179;  Indels  0;  Gaps  0;
```

RESULT 10

US-09-050-863-2

; Sequence 2, Application US/09050863

; Patent No. 6114111

; GENERAL INFORMATION:

;   APPLICANT: Lao, Ying

;   APPLICANT: Hiang, Betty

;   APPLICANT: Payan, Don

;   TITLE OF INVENTION: Mammalian Protein Interaction Cloning

;   TITLE OF INVENTION: System

;   NUMBER OF SEQUENCES: 5

;   CORRESPONDENCE ADDRESS:

;   ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert

;   STREET: 4 Embarcadero Center, Suite 3400

;   CITY: San Francisco

;   STATE: CA

;   COUNTRY: USA

;   ZIP: 94111-4187

;   COMPUTER READABLE FORM:

;   MEDIUM TYPE: Floppy disk

;   COMPUTER: IBM PC compatible

;   OPERATING SYSTEM: PC-DOS/MS-DOS

;   SOFTWARE: PatentIn Release #1.0, Version #1.30

;   CURRENT APPLICATION DATA:

;   APPLICATION NUMBER: US/09/050,863

;   FILING DATE: 30-MAR-1998

;   CLASSIFICATION:

;   ATTORNEY/AGENT INFORMATION:

;   NAME: Silva, Robin M.

;   REGISTRATION NUMBER: 38,304

;   REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS

;   TELECOMMUNICATION INFORMATION:

;   TELEPHONE: (415) 781-1989

;   TELEFAX: (415) 949-8711

; INFORMATION FOR SEQ ID NO: 2:

;   SEQUENCE CHARACTERISTICS:

;   LENGTH: 2580 base pairs

;   TYPE: nucleic acid

;   STRANDEDNESS: unknown

;   TOPOLOGY: unknown

;   MOLECULE TYPE: DNA

US-09-050-863-2

Query Match   9.5%; Score 50.6; DB 3; Length 2580;
Best Local Similarity 46.9%; Pred. No. 0.00075;
Matches 158; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

```
Qy    125 GGAGGACATGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACGGCCTGCCGA 184
          ||||||                 |||||||    ||||||||||||||||
Db    897 GGAGGAGGGGCAGGAGGCAGGAGCGGGGCCAGGAGGGGCAGGAGCAGGGGGCAGGA 956

Qy    185 TGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGAGCATCCGCAA 244
          ||   ||      ||||||
Db    957 GGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGA 1016

Qy    245 GTGGAGAGGAGGAGCAGAAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGACCGA 304
          ||   |||  ||||||
Db   1017 GGGGCAGGAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGA 1076

Qy    305 ACAGGAGTGCGCGGAGAAGGCCAAAAAAGACCTGGAACCAGCGCCAAAGTGA 364
          ||     ||||||
Db   1077 GCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGA 1136

Qy    365 ACAGGTTGAGAAGAACAAGATCAACAACAGGGCATCGGAGAGGGGCAGGAGGCTTTTGTGAAAGAATC 424
          ||||||
Db   1137 GGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGA 1196

Qy    425 CAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTG 461
          ||||||
Db   1197 GGAGGAGGGGCAGGAGGGGCAGGAGGAGCAGGAGGGGG 1233
```

RESULT 9

US-09-410-399-3

; Sequence 3, Application US/09410399

; Patent No. 6482587

; GENERAL INFORMATION:

;   APPLICANT: Robertson, Erle S.

;   APPLICANT: Cotter, Murray A.

;   TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA

;   TITLE OF INVENTION: to Genomic Host DNA

;   FILE REFERENCE: UM-03778

;   CURRENT APPLICATION NUMBER: US/09/410,399

;   CURRENT FILING DATE: 1999-10-01

;   NUMBER OF SEQ ID NOS: 6

;   SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

;   LENGTH: 1926

;   TYPE: DNA

;   ORGANISM: Epstein-Barr virus

US-09-410-399-3

Query Match   9.5%; Score 50.6; DB 3; Length 1926;
Best Local Similarity 46.9%; Pred. No. 0.00068;
Matches 158; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

```
Qy    125 GGAGGACATGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACGGCCTGCCGA 184
          ||||||
Db    514 GGAGGAGGGGCAGGAGGCAGGAGCGGGGCCAGGAGGGGCAGGAGCAGGGGGCAGGA 573

Qy    185 TGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGAGCATCCGCAA 244
          ||   ||      ||||||
Db    574 GGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGA 633

Qy    245 GTGGAGAGGAGGAGCAGAAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGACCGA 304
          ||   |||  ||||||
Db    634 GGGGCAGGAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGA 693

Qy    305 ACAGGAGTGCGCGGAGAAGGCCAAAAAAGACCTGGAACCAGCGCCAAAGTGA 364
          ||     ||||||
Db    694 GCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGA 753

Qy    365 ACAGGTTGAGAAGAACAAGATCAACAACAGGGCATCGGAGAGGGGCAGGAGGCTTTTGTGAAAGAATC 424
          ||||||
Db    754 GGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGA 813

Qy    425 CAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTG 461
          ||||||
Db    814 GGAGGAGGGGCAGGAGGGGCAGGAGGAGCAGGAGGGGG 850
```

RESULT 11
US-09-359-081-2
; Sequence 2, Application US/09359081
; Patent No. 6316223
; GENERAL INFORMATION:
;   APPLICANT: Lao, Ying
;   Hiang, Betty
;   Payan, Don
;   TITLE OF INVENTION: Mammalian Protein Interaction Cloning
;   System
;   NUMBER OF SEQUENCES: 5
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;   STREET: 4 Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94111-4187
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/359,081
;   FILING DATE: 22-Jul-1999
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 09/050,863
;   FILING DATE: <Unknown>
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Silva, Robin M.
;   REGISTRATION NUMBER: 38,304
;   REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 781-1989
;   TELEFAX: (415) 949-8711
;   INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2580 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: unknown
;   TOPOLOGY: unknown
;   MOLECULE TYPE: DNA
;   SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-359-081-2

Query Match   9.5%;   Score 50.6; DB 3;   Length 2580;
Best Local Similarity   46.9%;   Pred. No. 0.00075;
Matches 158;   Conservative   0;   Mismatches 179;   Indels   0;   Gaps   0;

Qy   125 GGAGGACATGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACGGGCCTGCCGA 184
          ||||| ||||||||||||||||| | ||||| || |||||||||||| ||||||
Db   897 GGAGGGGCAGGAGGGGCAGGAGCA GGAGGGGCAGGAGCCTGAGAGCATCCGCAA 956

Qy   185 TGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGGGGCATCCGCAA 244
          ||||| || | ||||||||||||||||| |||||||||||||||| |||||||||
Db   957 GGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCCTGAGAGCATCCGCAA 1016

Qy   245 GTGGAGAGGAGGAGCAGGAGAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGACCGA 304
          ||||||||||||||||||||||| | || ||||||||||||||||||||| |||||
Db   1017 GGGGCAGGAGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGA 1076

Qy   305 ACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAAGTGA 364
          |||||||| || ||||| | ||| ||||| ||||||||||||||| ||||||
Db   1077 GCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGA 1136

Qy   365 ACAGGTTGAGAAGAACAAGATCAACAACAGGGCATCGGAAGAGGCTTTTGTGAAAGAATC 424
          |||||||||||||| ||||||||| ||||||| |||||||| |||||
Db   1137 GGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCA 1196

Qy   425 CAAGGAGGAGACCCCAGGCACGAGTGGGAGAGGTG 461
          ||||||||||| |||||| ||||||||||||||||
Db   CAAGGAGGAGACCCCAGGCACGAGTGGGAGAGGTG

Db   1197 GGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGG 1233

RESULT 12
US-09-130-114-1/c
; Sequence 1, Application US/09130114
; Patent No. 5975807
; GENERAL INFORMATION:
;   APPLICANT: Horlick, Robert A.
;   Damaj, Bassam B.
;   Robbins, Alan K.
;   TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
;   TITLE OF INVENTION: From Multiple Transfected Episomes
;   FILE REFERENCE: 0867/1D903US1
;   CURRENT APPLICATION NUMBER: US/09/130,114
;   CURRENT FILING DATE: 1998-08-06
;   NUMBER OF SEQ ID NOS: 36
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 5452
;   TYPE: DNA
;   ORGANISM: VEBNA
US-09-130-114-1

Query Match   9.5%;   Score 50.6; DB 2;   Length 5452;
Best Local Similarity   46.9%;   Pred. No. 0.00097;
Matches 158;   Conservative   0;   Mismatches 179;   Indels   0;   Gaps   0;

Qy   125 GGAGGACATGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACGGGCCTGCCGA 184
          ||||| ||||||||||||||||| | |||| || |||||||||||| ||||||
Db   1908 GGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCCTGAGAGCATCCGCAA 1849

Qy   185 TGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGGGGCATCCGCAA 244
          ||||| || | ||||||||||||||||| |||||||||||||||| |||||||||
Db   1848 GGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGA 1789

Qy   245 GTGGAGAGGAGGAGCAGGAGAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGACCGA 304
          ||||||||||||||||||||||| | || ||||||||||||||||||||| |||||
Db   1788 GGGGCAGGAGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGA 1729

Qy   305 ACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAAGTGA 364
          |||||||| || ||||| | ||| ||||| ||||||||||||||| ||||||
Db   1728 GCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGA 1669

Qy   365 ACAGGTTGAGAAGAACAAGATCAACAACAGGGCATCGGAAGAGGCTTTTGTGAAAGAATC 424
          |||||||||||||| ||||||||| ||||||| |||||||| |||||
Db   1668 GGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCA 1609

Qy   425 CAAGGAGGAGACCCCAGGCACGAGTGGGAGAGGTG 461
          ||||||||||| |||||| ||||||||||||||||
Db   1608 GGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGG 1572

RESULT 13
US-09-647-344A-14/c
; Sequence 14, Application US/09647344A
; Patent No. 6586180
; GENERAL INFORMATION:
;   APPLICANT: Ruffner, Duane E.
;   Pierce, Michael L.
;   Chen, Zhidong
;   TITLE OF INVENTION: Directed Antisense Libraries
;   FILE REFERENCE: T6678.PCT.US
;   CURRENT APPLICATION NUMBER: US/09/647,344A
;   CURRENT FILING DATE: 2000-12-04
;   PRIOR APPLICATION NUMBER: PCT/US99/06742
;   PRIOR FILING DATE: 1999-03-28
;   NUMBER OF SEQ ID NOS: 50
; SEQ ID NO 14
;   LENGTH: 8705
;   TYPE: DNA
;   ORGANISM: Artificial Sequence

; FEATURE:
;, OTHER INFORMATION: pShuttle
US-09-647-344A-14

```
Query Match       9.5%;  Score 50.6;  DB 3;  Length 8705;
Best Local Similarity   46.9%;  Pred. No. 0.0011;
Matches 158;  Conservative  0;  Mismatches 179;  Indels  0;  Gaps  0;

Qy     125 GGAGGACATGGGACTACAGTCAATGGACGATGTGTTTCAGGAGGCTAACGGGCCTGCCGA 184
           |||||||         | |   | |        ||||||||||||||||||||||
Db    7773 GGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGA 7714

Qy     185 TGGCTACGCTGCGATTGCCCAGGCGACAGGTTGACTCAGGAGCCTGAGAGCATCCGCAA 244
           ||        |  | ||| |       |||||||||||||||||||||||||
Db    7713 GGGGCAGGAGGGGCAGGAGCCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGA 7654

Qy     245 GTGGAGAGAGGAGCAGAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGACCGA 304
           ||      | |   |           ||||||||||||||||||||||||
Db    7653 GGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGA 7594

Qy     305 ACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGAGGAGTGGAACCAGCGCCAAAGTGA 364
           || | |  |    |  || |         ||||||||||||||||||||
Db    7593 GCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGA 7534

Qy     365 ACAGGTTGAGAGAACAAGATCAACAACAGGGCATCGGAAGAGGCTTTTGTGAAAGAATC 424
           ||        |     |  | |   || |||||||||||||||
Db    7533 GGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCA 7474

Qy     425 CAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTG 461
           ||||||||||||||||
Db    7473 GGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGAGGGG 7437
```

RESULT 14
US-08-910-647-1
; Sequence 1, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION:
;, APPLICANT: Zuckermann et al.
;, TITLE OF INVENTION: Compositions and Methods for
;, TITLE OF INVENTION: Polynucleotide Delivery
;, NUMBER OF SEQUENCES: 4
;, CORRESPONDENCE ADDRESS:
;,   ADDRESSEE: Chiron Corporation
;,   STREET: 4560 Horton Street
;,   CITY: Emeryville
;,   STATE: California
;,   COUNTRY: U.S.A.
;,   ZIP: 94608-2916
;, COMPUTER READABLE FORM:
;,   MEDIUM TYPE: Floppy disk
;,   COMPUTER: IBM PC compatible
;,   OPERATING SYSTEM: PC-DOS/MS-DOS
;,   SOFTWARE: PatentIn Release #1.0, Version #1.30
;, CURRENT APPLICATION DATA:
;,   APPLICATION NUMBER: US/08/910,647
;,   FILING DATE:
;,   CLASSIFICATION: 514
;, ATTORNEY/AGENT INFORMATION:
;,   NAME: Fujita, Sharon M.
;,   REGISTRATION NUMBER: 38,459
;,   REFERENCE/DOCKET NUMBER: 1218.002
;, TELECOMMUNICATION INFORMATION:
;,   TELEPHONE: (510) 923-2706
;,   TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 1:
;, SEQUENCE CHARACTERISTICS:
;,   LENGTH: 9600 base pairs
;,   TYPE: nucleic acid
;,   STRANDEDNESS: single
;,   TOPOLOGY: linear
;, MOLECULE TYPE: DNA (genomic)
US-08-910-647-1

```
Query Match       9.5%;  Score 50.6;  DB 3;  Length 9600;
Best Local Similarity   46.9%;  Pred. No. 0.0012;
Matches 158;  Conservative  0;  Mismatches 179;  Indels  0;  Gaps  0;

Qy     125 GGAGGACATGGGACTACAGTCAATGGACGATGTGTTTCAGGAGGCTAACGGGCCTGCCGA 184
           |||||||         | |   | |        ||||||||||||||||||||||
Db     943 GGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGA 1002

Qy     185 TGGCTACGCTGCGATTGCCCAGGCGACAGGTTGACTCAGGAGCCTGAGAGCATCCGCAA 244
           ||        |  | ||| |       |||||||||||||||||||||||||
Db    1003 GGGGCAGGAGGGGCAGGAGCCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGA 1062

Qy     245 GTGGAGAGAGGAGCAGAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGACCGA 304
           ||      | |   |           ||||||||||||||||||||||||
Db    1063 GGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGA 1122

Qy     305 ACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGAGGAGTGGAACCAGCGCCAAAGTGA 364
           || | |  |    |  || |         ||||||||||||||||||||
Db    1123 GCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGA 1182

Qy     365 ACAGGTTGAGAGAACAAGATCAACAACAGGGCATCGGACGAGGCTTTTGTGAAAGAATC 424
           ||        |     |  | |   || |||||||||||||||
Db    1183 GGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCA 1242

Qy     425 CAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTG 461
           ||||||||||||||||
Db    1243 GGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGG 1279
```

RESULT 15
US-09-620-925-1
; Sequence 1, Application US/09620925
; Patent No. 6468986
; GENERAL INFORMATION:
;, APPLICANT: Zuckermann et al.
;, TITLE OF INVENTION: Compositions and Methods for
;, TITLE OF INVENTION: Polynucleotide Delivery
;, NUMBER OF SEQUENCES: 4
;, CORRESPONDENCE ADDRESS:
;,   ADDRESSEE: Chiron Corporation
;,   STREET: 4560 Horton Street
;,   CITY: Emeryville
;,   STATE: California
;,   COUNTRY: U.S.A.
;,   ZIP: 94608-2916
;, COMPUTER READABLE FORM:
;,   MEDIUM TYPE: Floppy disk
;,   COMPUTER: IBM PC compatible
;,   OPERATING SYSTEM: PC-DOS/MS-DOS
;,   SOFTWARE: PatentIn Release #1.0, Version #1.30
;, CURRENT APPLICATION DATA:
;,   APPLICATION NUMBER: US/09/620,925
;,   FILING DATE: 21-Jul-2000
;,   CLASSIFICATION: <Unknown>
;, PRIOR APPLICATION DATA:
;,   APPLICATION NUMBER: 08/910,647
;,   FILING DATE: <Unknown>
;, ATTORNEY/AGENT INFORMATION:
;,   NAME: Fujita, Sharon M.
;,   REGISTRATION NUMBER: 38,459
;,   REFERENCE/DOCKET NUMBER: 1218.002
;, TELECOMMUNICATION INFORMATION:
;,   TELEPHONE: (510) 923-2706
;,   TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 1:
;, SEQUENCE CHARACTERISTICS:
;,   LENGTH: 9600 base pairs
;,   TYPE: nucleic acid
;,   STRANDEDNESS: single
;,   TOPOLOGY: linear
;, MOLECULE TYPE: DNA (genomic)
;, SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-620-925-1

```
Query Match         9.5%;  Score 50.6;  DB 3;  Length 9600;
Best Local Similarity   46.9%;  Pred. No. 0.0012;
Matches 158;  Conservative  0;  Mismatches 179;  Indels  0;  Gaps  0;

Qy   125 GGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTGCCGA 184
             ||||||           ||      |          ||||||
Db   943 GGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGGGCAGGA 1002

Qy   185 TGGCTACCGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCCTGAGAGCATCCGCAA 244
              ||       ||      ||||||       |||||       |
Db  1003 GGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGGGCAGGAGGGGCAGGA 1062

Qy   245 GTGGAGAGAGGAGGCAGAAGAAAAGGCTGCAGGAGTTGCAGGATGCTGCCTCGAAGGTGACCGA 304
              |||||   |||         |||||||         |        ||
Db  1063 GGGGCAGGAGCAGCAGGGGGCAGGAGCAGGAGGGGCAGGAGCAGGGGCAGCGA 1122

Qy   305 ACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAAGTGA 364
             ||||||    |||         ||  |||||||||
Db  1123 GCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGGGCAGGA 1182

Qy   365 ACAGGTTGAGAAGAACAAGATCAACACAGGGCATCGGAAGAGGCTTTTGTGAAAGAATC 424
              ||||       |     |       |||||||      |||
Db  1183 GGGGCAGGAGCAGGAGGGGCGGCAGGAGCAGGAGGGGCAGGAGGGGGCAGGAGCA 1242

Qy   425 CAAGGAGGAGAGACCCCAGGCACAGAGTGGGAGAAGGTG 461
              |||||       |||  |||  ||||
Db  1243 GGGGGAGGGGGCAGGAGGGGCAGGAGCAGGAGGACGAGGAGGGGG 1279
```

Search completed: January 5, 2006, 21:35:23
Job time : 162 secs

```
            GenCore version 5.1.6
        Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:     January  5, 2006, 19:33:09 ; Search time 804 Seconds
                                 (without alignments)
                        5502.633 Million cell updates/sec

Title:         US-10-664-705-145
Perfect score: 535
Sequence:    1 cggccgccttcctggccag.........cgcctgcgctcggtgctcat 535

Scoring table: IDENTITY.NUC
            Gapop 10.0 , Gapext 1.0

Searched:      9793542 seqs, 413468905 residues

Total number of hits satisfying chosen parameters:      19587084


Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     Published Applications_NA_Main:*
            1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
            2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
            3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
            4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
            5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
            6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
            7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
            8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
            9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
           10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.


                              SUMMARIES
```

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	535	100.0	535	7	US-10-664-705-145	Sequence 145, App
2	471	88.0	982	3	US-09-917-800A-1407	Sequence 1407, Ap
3	471	88.0	982	6	US-10-388-934-227	Sequence 227, App
4	471	88.0	982	7	US-10-191-803-265	Sequence 265, App
5	471	88.0	982	7	US-10-152-319A-2036	Sequence 2036, Ap
6	451.8	84.4	1051	7	US-10-439-703-14	Sequence 14, Appl
7	451.8	84.4	1051	7	US-10-240-425-1304	Sequence 1304, Ap
8	410.8	76.8	636	9	US-10-956-157-3009	Sequence 3009, Ap
c 9	410.8	76.8	636	7	US-10-956-157-8244	Sequence 8244, Ap
10	387.8	72.5	1134	7	US-10-240-425-1303	Sequence 1303, Ap
c 11	380.8	71.2	894	3	US-09-823-245A-213	Sequence 213, App
c 12	379.2	70.9	907	3	US-09-823-245A-178	Sequence 178, App
13	364.4	68.1	490	3	US-09-918-995-20102	Sequence 20102, A
14	293.8	54.9	493	3	US-09-918-995-26855	Sequence 26855, A
15	268.8	50.2	485	3	US-09-918-995-27337	Sequence 27337, A
16	213.4	39.9	1164	6	US-10-264-049-328	Sequence 328, App
17	174.2	32.6	400	7	US-10-242-535A-10186	Sequence 10186, A
18	174.2	32.6	400	7	US-10-085-783A-10186	Sequence 10186, A
19	138.6	25.9	1105	3	US-09-880-107-2289	Sequence 2289, Ap
20	137.2	25.6	459	3	US-09-918-995-28203	Sequence 28203, A
21	125.8	23.5	338	7	US-10-242-535A-11137	Sequence 11137, A
22	125.8	23.5	338	7	US-10-085-783A-11137	Sequence 11137, A
23	122.2	22.8	419	7	US-10-242-535A-29408	Sequence 29408, A
24	122.2	22.8	419	7	US-10-085-783A-29408	Sequence 29408, A
25	115.4	21.6	473	3	US-09-918-995-10148	Sequence 10148, A
26	108	20.2	543	4	US-09-925-065A-555818	Sequence 555818,
27	104.6	19.6	489	7	US-10-424-599-116616	Sequence 116616,
28	90.2	16.9	612	4	US-09-925-065A-899680	Sequence 899680,
29	79.2	14.8	476	7	US-10-242-535A-24924	Sequence 24924, A
30	79.2	14.8	476	7	US-10-085-783A-24924	Sequence 24924, A
31	78.4	14.7	522	9	US-10-925-065A-642865	Sequence 642865,
32	78.2	14.6	1084	10	US-11-097-143-41372	Sequence 41372, A
33	78.2	14.6	3201	10	US-11-097-143-41371	Sequence 41371, A
34	72.6	13.6	484	9	US-09-918-995-20959	Sequence 20959, A
c 35	72	13.5	1764	9	US-10-450-763-25625	Sequence 25625, A
c 36	71	13.3	461	3	US-09-918-995-28043	Sequence 28043, A
37	66.6	12.4	411	4	US-10-242-535A-24460	Sequence 24460, A
38	66.6	12.4	411	7	US-10-085-783A-24460	Sequence 24460, A
c 39	65.2	12.2	658	3	US-09-871-161-327	Sequence 327, App
40	59.2	11.1	6227	6	US-10-240-485-51	Sequence 51, Appl
41	59	11.0	600	9	US-10-972-079-49308	Sequence 49308, A
42	57.2	10.7	439	3	US-09-918-995-10020	Sequence 10020, A
43	57	10.7	628	6	US-10-029-386-22859	Sequence 22859, A
c 44	57	10.7	3920	10	US-11-097-143-21013	Sequence 21013, A
c 45	55.8	10.4	659158	3	US-09-771-208-20	Sequence 20, Appl

```
                              ALIGNMENTS


RESULT 1
US-10-664-705-145
; Sequence 145, Application US/1664705
; Publication No. US20040152107A1
; GENERAL INFORMATION:
;  APPLICANT: Altar, Anthony C.
;  APPLICANT: Laeng, Pascal
;  APPLICANT: Young, Theresa A.
;  APPLICANT: Charles, Vinod
;  APPLICANT: Bukhman, Yury
;  APPLICANT: Jurata, Linda
;  TITLE OF INVENTION: GENE SIGNATURE OF ELECTROSHOCK THERAPY AND METHODS OF USE
;  FILE REFERENCE: 03235/100M087-US2
;  CURRENT APPLICATION NUMBER: US/10/664,705
;  CURRENT FILING DATE:  2003-09-18
;  PRIOR APPLICATION NUMBER: US 60/411,718
;  PRIOR FILING DATE:  2002-09-18
;  PRIOR APPLICATION NUMBER: US 60/431,882
;  PRIOR FILING DATE:  2002-09-18
;  PRIOR APPLICATION NUMBER: US 60/479,970
;  PRIOR FILING DATE:  2003-06-18
;  NUMBER OF SEQ ID NOS: 152
;  SOFTWARE: Patentin version 3.1
; SEQ ID NO 145
;  LENGTH: 535
;  TYPE: DNA
;  ORGANISM: Rattus norvegicus
US-10-664-705-145

Query Match             100.0%;  Score 535;  DB 7;  Length 535;
Best Local Similarity   100.0%;  Pred. No. 1.2e-144;
Matches  535;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CGGCCGCCTTCCTGGCCAGCAGGAGGACTACAGTGCTGGCATCGAGAATGACTCGGGTT   60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CGGCCGCCTTCCTGGCCAGCAGGAGGACTACAGTGCTGGCATCGAGAATGACTCGGGTT   60

Qy         61 TCGGGGCACCTGCCGCCGCCAGGTGCCTCTGCGCAGCCCGGACTCGGCGAGCCGGGGGTG  120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCGGGGCACCTGCCGCCGCCAGGTGCCTCTGCGCAGCCCGGACTCGGCGAGCCGGGGGTG  120

Qy        121 GTTCGGAGGACATGGGACTACACAGTCAATGACGATGAGTGTTTCAGGAGGCTAACGGGCTG  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTTCGGAGGACATGGGACTACACAGTCAATGACGATGAGTGTTTCAGGAGGCTAACGGGCCTG  180

Qy        181 CCGATGCTACGCTCGGATTGCCCAGGCGGCGACAGGTTGACTCAGGAGAGCATCC  240
```

```
Db   181 CCGATGGCTACGCTGCGATTGCCCAGCGCGGACAGGTTGACTTCAGGAGCCTGAGAGCATCC 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   241 GCAAGTGGAGAGGAGGAGGAGAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCAAGTGGAGAGGAGGAGGAGAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA 300
Qy   301 CCGAACAGGAGTGCGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CCGAACAGGAGTGCGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 360
Qy   361 GTGAACAGGTTGAGAAGAACAAGATCAACACAGGGCATCCGAAAGAGGCTTTTGTGAAAG 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GTGAACAGGTTGAGAAGAACAAGATCAACACAGGGCATCCGAAAGAGGCTTTTGTGAAAG 420
Qy   421 AATCCAAGGAGGAGACCCAGCCACAGAGTGGAGAGGTGGCCCAGCTGTGTGACTTCA 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 AATCCAAGGAGGAGACCCAGCCACAGAGTGGAGAGGTGGCCCAGCTGTGTGACTTCA 480
Qy   481 ACCCTAAGCAGCAGCAATGTAAAGACGTGTCCCGCTGCGTCGGTGCTCAT 535
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ACCCTAAGCAGCAGCAATGTAAAGACGTGTCCCGCTGCGTCGGTGCTCAT 535
```

RESULT 2
US-09-917-800A-1407
; Sequence 1407, Application US/09917800A
; Patent No. US2002011946[2]A1
; GENERAL INFORMATION:
;   APPLICANT: Mendrick, Donna
;   APPLICANT: Porter, Mark
;   APPLICANT: Johnson, Kory
;   APPLICANT: Castle, Arthur
;   APPLICANT: Elashoff, Michael
;   APPLICANT: Gene Logic, Inc.
;   TITLE OF INVENTION: Molecular Toxicology Modeling
;   FILE REFERENCE: 44921-5038-US
;   CURRENT APPLICATION NUMBER: US/09/917,800A
;   CURRENT FILING DATE: 2001-07-31
;   PRIOR APPLICATION NUMBER: US 60/222,040
;   PRIOR FILING DATE: 2000-07-31
;   PRIOR APPLICATION NUMBER: US 60/222,880
;   PRIOR FILING DATE: 2000-11-02
;   PRIOR APPLICATION NUMBER: US 60/290,029
;   PRIOR FILING DATE: 2001-05-11
;   PRIOR APPLICATION NUMBER: US 60/290,645
;   PRIOR FILING DATE: 2001-05-15
;   PRIOR APPLICATION NUMBER: US 60/292,336
;   PRIOR FILING DATE: 2001-05-22
;   PRIOR APPLICATION NUMBER: US 60/295,798
;   PRIOR FILING DATE: 2001-06-06
;   PRIOR APPLICATION NUMBER: US 60/297,457
;   PRIOR FILING DATE: 2001-06-13
;   PRIOR APPLICATION NUMBER: US 60/298,884
;   PRIOR FILING DATE: 2001-06-19
;   PRIOR APPLICATION NUMBER: US 60/303,459
;   PRIOR FILING DATE: 2001-07-09
;   SOFTWARE: PatentIn Ver. 2.1
;   NUMBER OF SEQ ID NOS: 1740
;   SEQ ID NO 1407
;   LENGTH: 982
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US2002011946[2]A1 M15883
US-09-917-800A-1407

Query Match   88.0%; Score 471; DB 3; Length 982;
Best Local Similarity   90.8%; Pred. No. 4.7e-126;
Matches 535; Conservative   0; Mismatches   0; Indels 54; Gaps   1;

```
Qy    1 CGGCCGCCTTCCTGGCCCAGCAGGAGGAGCAGGAGGATTGCTGGCATCGAGAATGACTCGGGGTT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   99 CGGCCGCCTTCCTGGCCCAGCAGGAGGAGCAGGAGGATTGCTGGCATCGAGAATGACTCGGGGTT 158
Qy   61 TCGGGGCACCTGCGCCAGCCAGGTGTCGCAGGAGTTGGATGCTGCCTCGAAGGTG 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  159 TCGGGGCACCTGCGCCAGCCAGGTGCCTCGCAGCCCGGACTGCCGCCCGCGGGGTG 218
Qy  121 GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTCAGGAGGCTAACGGCCTG 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  219 GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTCAGGAGGCTAACGGCCTG 278
Qy  181 CCGATGGCTACGCTGCGATTGCCCAGCGCGGACAGGTTGACTCAGGAGCCTGAGAGCATCC 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  279 CCGATGGCTACGCTGCGATTGCCCAGCGCCGGACAGGTTGACTCAGGAGCCTGAGAGCATCC 338
Qy  241 GCAAGTGGAGAGGAGGAGCAGGAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  339 GCAAGTGAGAGAGAGAGGAGCAGAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA 398
Qy  301 CCGAACAGGAGTGCGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  399 CCGAACAGGAGTGCGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 458
Qy  361 GTGAACAGGTTGAGAAGAACAAGATCAACAACAG------------------------ 394
        |||||||||||||||||||||||||||||||||
Db  459 GTGAACAGGTTGAGAAGAACAAGATCAACAGGATCGCTGCACAAAGCGTTCTACCAGC 518
Qy  395 ---------------------------GGCATCCGGAAGAGGCTTTTGTGAAAGAATCCA 426
                                   |||||||||||||||||||||||||||||||||
Db  519 AGCCAGATGCTGATACCATTGGCTATGTGGCATCCGGAAGAGGCTTTTGTGAAAGAATCCA 578
Qy  427 AGAGGAGACCCAGCCACAGAGTGGGAGAGGTGGCCCAGCTGTGTGACTTCAACCCTA 486
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  579 AGAGGAGACCCAGCCACAGAGTGGGAGAGGTGGCCCAGCTGTGTGACTTCAACCCTA 638
Qy  487 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCTGCGTCGGTGCTCAT 535
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  639 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCTGCGTCGGTGCTCAT 687
```

RESULT 3
US-10-388-934-227
; Sequence 227, Application US/10388934
; Publication No. US2004000554[7]A1
; GENERAL INFORMATION:
;   APPLICANT: Boess, Franziska
;   APPLICANT: Suter-Dick, Laura
;   APPLICANT: Wolf, Detlef
;   TITLE OF INVENTION: BIOMARKERS AND EXPRESSION PROFILES FOR TOXICOLOGY
;   FILE REFERENCE: 21199
;   CURRENT APPLICATION NUMBER: US/10/388,934
;   CURRENT FILING DATE: 2003-03-14
;   PRIOR APPLICATION NUMBER: 02005336.9
;   PRIOR FILING DATE: 2002-03-14
;   PRIOR APPLICATION NUMBER: 02015657.6
;   PRIOR FILING DATE: 2002-07-17
;   NUMBER OF SEQ ID NOS: 862
;   SOFTWARE: PatentIn version 3.1
;   SEQ ID NO 227
;   LENGTH: 982
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus (No. US2004000554[7]Alway rat)
US-10-388-934-227

Query Match   88.0%; Score 471; DB 6; Length 982;
Best Local Similarity   90.8%; Pred. No. 4.7e-126;
Matches 535; Conservative   0; Mismatches   0; Indels 54; Gaps   1;

```
Qy    1 CGGCCGCCTTCCTGGCCCAGCAGGAGGAGCAGGAGGATTGCTGGCATCGAGAATGACTCGGGGTT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   99 CGGCCGCCTTCCTGGCCCAGCAGGAGGAGCAGGAGGATTGCTGGCATCGAGAATGACTCGGGGTT 158
Qy   61 TCGGGGCACCTGCGCCAGCCAGGTGCCTCGCAGCCCGGACTGCCGCCCGCGGGGTG 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  159 TCGGGGCACCTGCGCCAGCCAGGTGCCTCGCAGCCCGGACTGCCGCCCGCGGGGTG 218
```

Db   99 CGGCCGCCTTCCTGGCCACCTGCCAGCGAGAGAGCGAGAATTGCTGCGCACCCGGACTCGCGAGCTCGCGAGCGGGGGTT 158
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   61 TCGGGGCACCTGCCAGCTGGCCTCTGCGCACCCGGACTCGCGAGGCTAACGGGCCTG 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 159 TCGGGGCACCTGCCAGCTGGCCTCTGCGCACCCGGACTCGCGAGGCTAACGGGCCTG 218

Qy 121 GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTGAGAGCATCC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 219 GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTGAGAGCATCC 278

Qy 181 CCGATGGCTACGCTGCGATTGCCCAGGCGACAGGTTGACTCAGGAGTGTCGCCTGAAGGTGA 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 279 CCGATGGCTACGCTGCGATTGCCCAGGCGACAGGTTGACTCAGGAGTGTCGCCTGAAGGTGA 338

Qy 241 GCAAGTGGAGAGGAGCAGAGAAAAAGCTGCGAGGAGTTGGATGCTGCGCTCGAAGGTGA 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 339 GCAAGTGGAGAGGAGCAGAGAAAAAGCTGCGAGGAGTTGGATGCTGCGCTCGAAGGTGA 398

Qy 301 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 399 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA 458

Qy 361 GTGAACAGGTTGAGAGACAAGATCAACAACAG------------------------ 394
        ||||||||||||||||||||||||||||||||
Db 459 GTGAACAGGTTGAGAAGAACAAGATCAACAACAGATCGCTGACAAAGCGTTCTACCAGC 518

Qy 395 -----------------------------GGCATCGGAAGAGGCTTTTGTGAAAGAATCCA 426
                                      |||||||||||||||||||||||||||||||
Db 519 AGCCAGATGCTGATACCATTGGCTATGTGGCATCGGAAGAGGCTTTTGTGAAAGAATCCA 578

Qy 427 AGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTA 486
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 579 AGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTA 638

Qy 487 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTGCCGCTCGCGGTGCTCAT 535
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 639 AGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTGCCGCTCGCGGTGCTCAT 687

RESULT 4
US-10-191-803-265
; Sequence 265, Application US/10191803
; Publication No. US20040014040A1
GENERAL INFORMATION:
;   APPLICANT: MENDRICK, Donna
;   APPLICANT: PORTER, Mark
;   APPLICANT: JOHNSON, Kory
;   APPLICANT: HIGGS, Brandon
;   APPLICANT: CASTLE, Arthur
;   APPLICANT: ELASHOFF, Michael
;   TITLE OF INVENTION: Cardiotoxin Molecular Toxicology Modeling
;   FILE REFERENCE: 44921-5090US
;   CURRENT APPLICATION NUMBER: US/10/191,803
;   CURRENT FILING DATE: 2002-07-10
;   PRIOR APPLICATION NUMBER: US 60/303,819
;   PRIOR FILING DATE: 2001-07-10
;   PRIOR APPLICATION NUMBER: US 60/305,623
;   PRIOR FILING DATE: 2001-07-17
;   PRIOR APPLICATION NUMBER: US 60/369,351
;   PRIOR FILING DATE: 2002-04-03
;   PRIOR APPLICATION NUMBER: US 60/377,611
;   PRIOR FILING DATE: 2002-05-06
;   NUMBER OF SEQ ID NOS: 1140
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 265
;   LENGTH: 982
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20040014040A1 NM_053835
US-10-191-803-265

Query Match   88.0%; Score 471; DB 7; Length 982;
Best Local Similarity 90.8%; Pred. No. 4.7e-126;
Matches 535; Conservative 0; Mismatches 54; Indels 0; Gaps 1;

Qy   1 CGGCCGCCTTCCTGGCCCAGCAGGAGGAGGAGCGAGAATTGCTGCGCATCGGCAATGACTCGCGGGTT 60

RESULT 5
US-10-152-319A-2036
; Sequence 2036, Application US/10152319A
; Publication No. US20040072160A1
GENERAL INFORMATION:
;   APPLICANT: Mendrick, Donna
;   APPLICANT: Porter, Mark
;   APPLICANT: Johnson, Kory
;   APPLICANT: Higgs, Brandon
;   APPLICANT: Castle, Arthur
;   APPLICANT: Elashoff, Michael
;   TITLE OF INVENTION: Molecular Toxicology Modeling
;   FILE REFERENCE: 44921-5089-US
;   CURRENT APPLICATION NUMBER: US/10/152,319A
;   CURRENT FILING DATE: 2002-05-22
;   PRIOR APPLICATION NUMBER: US 60/292,335
;   PRIOR FILING DATE: 2001-05-22
;   PRIOR APPLICATION NUMBER: US 60/297,523
;   PRIOR FILING DATE: 2001-06-13
;   PRIOR APPLICATION NUMBER: US 60/298,925
;   PRIOR FILING DATE: 2001-06-19
;   PRIOR APPLICATION NUMBER: US 60/303,810
;   PRIOR FILING DATE: 2001-07-10
;   PRIOR APPLICATION NUMBER: US 60/303,807
;   PRIOR FILING DATE: 2001-07-10
;   PRIOR APPLICATION NUMBER: US 60/303,808
;   PRIOR FILING DATE: 2001-07-10
;   PRIOR APPLICATION NUMBER: US 60/315,047
;   PRIOR FILING DATE: 2001-08-28
;   PRIOR APPLICATION NUMBER: US 60/324,928
;   PRIOR FILING DATE: 2001-09-27
;   PRIOR APPLICATION NUMBER: US 60/330,867
;   PRIOR FILING DATE: 2001-11-01

; PRIOR APPLICATION NUMBER: US 60/330,462
; PRIOR FILING DATE: 2001-10-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2221
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2036
; LENGTH: 982
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. NM_053835
US-10-152-319A-2036

Query Match   88.0%;   Score 471;   DB 7;   Length 982;
Best Local Similarity   90.8%;   Pred. No. 4.7e-126;
Matches   535;   Conservative   0;   Mismatches   0;   Indels   54;   Gaps   1;

```
Qy      1 CGGCCGCCTTCTGGCCCAGCAGGAGAGCGAGATTGCTGGCATGCAGAATGACTCGGGTT  60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     99 CGGCCGCCTTCTGGCCCAGCAGGAGAGCGAGATTGCTGGCATGCAGAATGACTCGGGTT 158

Qy     61 TCGGGGCACCTGCCGCGGCAGGTGGCCTCTGCGCAGCCCGCGGGGTG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    159 TCGGGGCACCTGCCGCGGCAGGTGGCCTCTGCGCAGCCCGCGGGGTG 218

Qy    121 GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219 GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG 278

Qy    181 CCGATGGCTACGGTGCGATTGCCCAGGCGGCAGGTTGACTCAGGAGCATCC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    279 CCGATGGCTACGGTGCGATTGCCCAGGCGGCAGGTTGACTCAGGAGCATCC 338

Qy    241 GCAAGTGAGAGTGGCGGGAGAGGCCAAAAAAGACCTGGAGGTGAACCAGG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 GCAAGTGAGAGTGGCGGGAGAGGCCAAAAAAGACCTGGAGGTGAACCAGG 398

Qy    301 CCGAACAGGAGTGGCGGGAGAGGCCAAAAAAGACCTGGAGGTGAACCAGG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    399 CCGAACAGGAGTGGCGGGAGAGGCCAAAAAAGACCTGGAGGTGAACCAGG 458

Qy    361 GTGAACAGGTTGAAGAACAAGATCAACAACAG---------- 394
          ||||||||||||||||||||||||||||||||
Db    459 GTGAACAGGTTGAAGAACAAGATCAACAACAGGATCGTTGAAAGCGTTCTACCAGC 518

Qy    395 -------------GGCATCGGAAGAGGCTTTTGTGAAAGAATCCA 426
                       |||||||||||||||||||||||||||||||
Db    519 AGCCAGATGCTGATACCATTGGCATCGGAAGAGGCTTTTGTGAAAGAATCCA 578

Qy    427 AGGAGGAGACCCCCAGGCACAGAGTGGGAGAAGGTGGCCAGCTGTGTCCCGCCTA 486
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    579 AGGAGGAGACCCCCAGGCACAGAGTGGGAGAAGGTGGCCAGCTGTGTCCCGCCTA 638

Qy    487 AGAGCAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTCGGTGCTCAT 535
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    639 AGAGCAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTCGGTGCTCAT 687
```

RESULT 6
US-10-439-703-14
; Sequence 14, Application US/10439703
; Publication No. US20040018527A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Jenny
; APPLICANT: O'Connell, Peter
; TITLE OF INVENTION: Differential Patterns of Gene Expression that Predict for Docetax
; TITLE OF INVENTION: Chemosensitivy and Chemoresistance
; FILE REFERENCE: HO-P2482US1/10205813
; CURRENT APPLICATION NUMBER: US/10/439,703
; CURRENT FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: US 60/381,141
; PRIOR FILING DATE: 2002-05-17
; NUMBER OF SEQ ID NOS: 91

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1051
; TYPE: DNA
; ORGANISM: Human
US-10-439-703-14

Query Match   84.4%;   Score 451.8;   DB 7;   Length 1051;
Best Local Similarity   90.3%;   Pred. No. 1.7e-120;
Matches   483;   Conservative   0;   Mismatches   52;   Indels   0;   Gaps   0;

```
Qy      1 CGGCCGCCTTCTGGCCCAGCAGGAGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT  60
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    210 CGGCCGCCTTCTGGCCCAGCAGGAGAGCGAGATTGCAGGCATAGAGAACGACGAGGGCT 269

Qy     61 TCGGGGCACCTGCCGCGGCAGGTGGCCTCTGCGCAGCCCGAGCTCGGCGAGCCGGGGCTG 120
          ||||||||||||||||||||||||| ||||||||||| ||| |||||||||||||| |||
Db    270 TCGGGGCACCTGCCGCGGCAGCCATGCGGCCGCCCCGCAGCCGGGGCCCCAGAGTGGGGCTG 329

Qy    121 GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG 180
          |||||||||||||||||||||||| ||||||| ||||||||||||||||||||||||||
Db    330 GTTCTGAGGACATGGGGACCACAGTCAATGGAGATGTGTTTCAGGAGGCCAACGCGTCCTG 389

Qy    181 CCGATGGCTACGGTGCGATTGCCCAGGCGGCAGGTTGACTCAGGAGCATCC 240
          |||||||||||| ||| ||| |||||||||||||||||||||||||||||
Db    390 CTGATGGCTACGGTGCCAGCCATTGCCCAGGCTGCACAGGCTGACCCCAGGAGCATCC 449

Qy    241 GCAAGTGAGAGTGGCGGGAGAGGCCAAAAAAGACCTGGAGGTGAGAAGGTGA 300
          ||||||||||||||| ||||||||||||||| |||||| ||||||||||
Db    450 GCAAGTGCGAGAGGAGGCAAGAGGAGAAACCGCTGCAAGAGCTGGATTGCTGCATCTAAGGTCA 509

Qy    301 CCGAACAGGAGTGGCGGGAGAGGCCAAAAAAGACCAGCGCCAA 360
          ||||||||||||||| ||||||||||||| ||||||||||
Db    510 CCGAACAGGACGGAATGCGCGGAGAAGGACCTGGAGGTGGAACCAGCGCCAGA 569

Qy    361 GTGAACAGGTTGAAGAACAACAGGCATCGGAAGAGGCTTTTGTGAAAG 420
          |||||||||||||||||| ||||||||| |||||||||||||||||
Db    570 GTGAACAAGTAGAAGAAGCAAGATCAACACCCGGAGGAGGCCCGAGGAGGCTTTCGTGAAGG 629

Qy    421 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCAGC 480
          ||||||||||||||||||| ||||||||||||||||||||||||||
Db    630 AATCCAAGGAGGAGACCCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTATGTCGATTCA 689

Qy    481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTCGGTGCTCAT 535
          |||  |||||||||||||||||||||||||||||||||||||||||||||
Db    690 ACCCCAAGCAGCAGCAAGCAGTGCAAAAGATGTGTCCCGCCTCGGCGGTGTCTCAT 744
```

RESULT 7
US-10-240-425-1304
; Sequence 1304, Application US/10240425
; Publication No. US2004003502A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Amanda
; APPLICANT: Boland, Joseph F.
; APPLICANT: Lord, Reginald V.
; APPLICANT: Alvarez, Chris
; APPLICANT: Wetzel, Jon C.
; APPLICANT: Scherf, Uwe
; APPLICANT: Vockley, Joseph G.
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
; FILE REFERENCE: 44921-5026
; CURRENT APPLICATION NUMBER: US/10/240,425
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: PCT/US01/09847
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/193,446
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 1588
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1304
; LENGTH: 1051
; TYPE: DNA

```
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20040033502A1 M20470
US-10-240-425-1304

  Query Match          84.4%;  Score 451.8;  DB 7;   Length 1051;
  Best Local Similarity  90.3%;  Pred. No. 1.7e-120;
  Matches  483;  Conservative  0;  Mismatches  52;  Indels  0;  Gaps  0;

Qy     1  CGGCCGCCTTCCTGGCCCAGCGAGGCGAGATTGCTGGCATCGAGAATGACTCGGGTT   60
          |||||| ||||||||||||||||| |||||  ||||| ||||||| |||||||||||||
Db   210  CGGCCGCCTTCCTGGCCCAGCCCAGCAGGAGCGAGATTGCAGGCATCGACGAGAGGGCT  269

Qy    61  TCGGGGCACCTGCCGCCGCCAGCCAGGTGGCCTTCGCGCAGCCCGACTCGCGAGCGGGGTG  120
          |||||||||||||||||| ||||||||||||||||||||||| |||||||||||||| |||
Db   270  TCGGGGCACCTGCCGCCGCCAGCAGCCAGCCATGCGGCCCCGCAGCCCGACCAGTGGGGCTG  329

Qy   121  GTTCGGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGGCCTG  180
          |||| ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   330  GTTCTGAGGACATGGGGACCACAGTCAATGGAGATGTGTTTCAGGAGGCCAACGGGTCCTG  389

Qy   181  CCGATGGCTACGCTGCGATTGCCCAGGCCGACAGGTTGACTCAGGAGCCTGAGAGCATCC  240
          |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db   390  CTGATGGCTACGCTGCGATTGCCCAGGCTGACACGCTGACAGGAGCCTGAGAGCATCC  449

Qy   241  GCAAGTGAGAGAGTGGCGGGAGGAAGGCAGAGAGAAAAAGACCTGGAACCAGCGCCAAA  300
          |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db   450  GCAAGTGAGAGAGTGGCGGGAGGAAGGCAGAGAGGAGATGCTGCCTCGGAAGGTGA  509

Qy   301  CCGAACAGAGTGGCGGGAGGAAGGCCAAAAAAGACCTGGAACCAGCGCCAAA  360
          |||||||||||||||||||||||||||||||||| ||||||||||||||
Db   510  CGGAACAGAGTGGCGGGAGGAAGGCCAAGAATGGCGGGAGGAAGGCCAAG  569

Qy   361  GTGAACAGAGTTGAGAGAACAAGATCAACACAGGGCATCGGAAGAGGCTTTTGTGAAAG  420
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||| |||
Db   570  GTGAACAGAGTAGAGAGAACAAGATCAACAACCGGGCATCCGAGGAGGCTTTCGTGAAGG  629

Qy   421  AATCCAAGGAGGAGAGACCCCCAGGCACAGAGTGGGAGAAGTGGCCCAGCTGTGTGACTTCA  480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   630  AATCCAAGGAGGAGAGACCCCCAGGCACAGAGTGGGAGAAGTGGCCCAGCTATGTGACTTCA  689

Qy   481  ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCTCGGTCGGTGCTCAT  535
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   690  ACCCCAAGAGCAGCAAGCAGTGCAAAGATGTGTCCCGCTCGGTCGGTGCTCAT  744
```

RESULT 8
US-10-956-157-3009
;   Sequence 3009, Application US/10956157
;   Publication No. US20050118625A1
;   GENERAL INFORMATION:
;   APPLICANT: Wyeth
;   APPLICANT: Mounts, William
;   TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
;   TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
;   FILE REFERENCE: 031896-043000 (AM 101081)
;   CURRENT APPLICATION NUMBER: US/10/956,157
;   CURRENT FILING DATE: 2004-10-04
;   NUMBER OF SEQ ID NOS: 319805
;   SOFTWARE: PatentIn version 3.2
;   SEQ ID NO 3009
;   LENGTH: 636
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-956-157-3009

  Query Match   76.8%; Score 410.8; DB 9;   Length 636;
  Best Local Similarity 89.5%; Pred. No. 1.2e-108;
  Matches 453; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

```
Qy    31  AGATTGCTGGCATCGAGAATGACTCGGGTTCGGGGCACCTGCCGCCGCCAGGTGGCCT   90
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db     1  AGATTGCAGGCATAGAGAACGACGAGAGGGCTTCGGGGCACCTGCCGCCGCCAGCAGCC   60
```

RESULT 9
US-10-956-157-8244/c
;   Sequence 8244, Application US/10956157
;   Publication No. US20050118625A1
;   GENERAL INFORMATION:
;   APPLICANT: Wyeth
;   APPLICANT: Mounts, William
;   TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
;   TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
;   FILE REFERENCE: 031896-043000 (AM 101081)
;   CURRENT APPLICATION NUMBER: US/10/956,157
;   CURRENT FILING DATE: 2004-10-04
;   NUMBER OF SEQ ID NOS: 319805
;   SOFTWARE: PatentIn version 3.2
;   SEQ ID NO 8244
;   LENGTH: 636
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-956-157-8244

  Query Match   76.8%; Score 410.8; DB 9;   Length 636;
  Best Local Similarity 89.5%; Pred. No. 1.2e-108;
  Matches 453; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

```
Qy    31  AGATTGCTGGCATCGAGAATGACTCGGGTTCGGGGCACCTGCCGCCGCCAGGTGGCCT   90
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   636  AGATTGCAGGCATAGAGAACGACGAGAGGGCTTCGGGGCACCTGCCGCCGCCAGCAGCC  577

Qy    91  CTGCGCAGCCCGACTCGCGAGCGGGGTGGTTC-GGAGGACATGGGGACTACAGTCAAT  149
          |||||||||||||||||| |||||||||||||  ||||||||||||||||| |||||||
Db   576  CCGCGCAGCCCGACCAGTGGGGCTGGTTCTTGGAGGACATGGGGACCACAGTCAAT  517

Qy   150  GGAGATGTGTTTCAGGAGGCTAACGGGCTACGGCTACGCTGCGATTGCCCAGGCG  209
          ||||||||||||||||||||| ||||||||||| |||||||||||||||||||||
Db   516  GGAGATGTGTTTCAGGAGGCCAACGGGTCCTGCTGATGGCTACGCTGCGATTGCCCAGGCT  457

Qy   210  GACAGGTTGACTCAGGAGCCTGAGAGCATCAAGATCAAC  269
          |||||||| |||||||||||||||||||||| ||||| |
Db   456  GACAGGCTGACCCAGGAGCCTGAGAGCATCAAGATCAAC  397
```

```
Qy    91  CTGCGCAGCCCGACTCGCGAGCGGGGTGGTTC-GGAGACATGGGGACTACAGTCAAT  149
          |||||||||||||||||| ||||||||||||||  |||| |||||||||| ||||||||
Db    61  CCGCGCAGCCCGGCCCCACAGTGGGCTGGTTCTTCTTGAGGACACATGGGACATGGGGACCACAGTCAAT  120

Qy   150  GGAGATGTGTTTCAGGAGGCTAACGGGCTACGGCTACGCTGCGATTGCCCAGGCG  209
          ||||||||||||||||||||| ||||||||||| ||||||||||||||||||||||
Db   121  GGAGATGTGTTTCAGGAGGCCAACGGCCATTGCCGTGATGGCTACGCTGCGATTGCCCAGGCT  180

Qy   210  GACAGGTTGACTCAGGAGCATCCGAAGTGACATGGGAACAGAGGAACAGATCAAC  269
          |||||||| |||||||||||||| |||||| |||||||||||| ||||| |
Db   181  GACAGGCTGACCCAGGAGCATCCGGAGTGACATGGGGAACAGAGGAGAACAGATCAAC  240

Qy   270  CTGCAGGAGTGGATGCTGCCTCGGAAGGTGACGAACAGAGGAAGGCCAAA  329
          ||||||||| |||||||||||||| |||| |||||||||||| |||||
Db   241  CTGCAAGAGCTGGATGCTGCCTAAGGTTCACGGAACAGAGGAGAACAAGGCCAAG  300

Qy   330  AAAGACCTGGAACCAGCCGAAGAGTGAACAGAGGTGAACAAGATCAAC  389
          |||||||||||||| ||||||||||||||| ||||||||| |||||||
Db   301  AAGGACCTGAGGAGTGGAACCAGCCCAGTGAACAGAGTGGAACAAGATCAAC  360

Qy   390  AACAGGGCATCGGAAGAGGCTTTTGTGAAAGAATCCAAGGGAGCACAGAG  449
          ||||| ||||| |||||||||| |||||| ||||||||||| |||||||
Db   361  AACCGGGCATCCGAGGAGGCTTTCGTGAAGGAATCCAAGGAGGAGCACAGAG  420

Qy   450  TGGGAAGGTGGCCCAGTCGTGTGACTTCAACCCTAAGAGCAGCAAGCAATGTAAAGAC  509
          ||||||||| ||||| ||||||||||||||||| |||||||||||||||| ||||||
Db   421  TGGGAAGGTGGCCCAGCTATGTGACTTCAACCCCAAGAGCAGCAAGCAGTGCAAAGAT  480

Qy   510  GTGTCCCGCTCGGTCGGTGCTCAT  535
          |||||||||||||||||||||||||
Db   481  GTGTCCCGCTCGGTCGGTGCTCAT  506
```

```
Qy    270  CTGCAGGAGTTGGATGCTGCCTCGAAGTGCAGCAGGAGAGTGGCGGGAGAAGGCCAAA  329
           |||||| |||||||||||||||||||||||||||||||||||| |||||||||||||||
Db    396  CTGCAAGAGCTGATGCTGCCTCGCATCCAGAATGACTCGGGTT||||||||||||||  337

Qy    330  AAAGACCTGGAGGAGTGGAACCAGCGCCAAAGTGAACAGTTGAGAGAAGAACAAGATCAAC  389
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    336  AAGGACCTGGAGGAGTGGAACCAGCGCCAGAGTGAACAAGTAGAAGAAGAACAAGATCAAC  277

Qy    390  AACAGGGCATCGGAAGAGGCTTTTGTGAAAGAATCCAAGGAGGAGACCCCAGGCACAGAG  449
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    276  AACCGGGCATCCGAGGAGGCTTTCGTGAAGGAATCCAAGGAGGAGCAAGCACCCAGCAGAG  217

Qy    450  TGGGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTAAGAGAGCAAGCAATGTAAAGAC  509
           |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db    216  TGGGAGAAGGTGGCCCAGCTGTGACTTCAACCCCAGAGAGCAAGCAGTGCAGTAAAGAT  157

Qy    510  GTGTCCCGCCTGCGCTCGGTGCTCAT  535
           ||||||||||||||||||||||||||
Db    156  GTGTCCCGCCTGCGCTCGGTGCTCAT  131
```

```
RESULT 10
US-10-240-425-1303
; Sequence 1303, Application US/10240425
; Publication No. US20040033502A1
; GENERAL INFORMATION:
;    APPLICANT: Williams, Amanda
;    APPLICANT: Boland, Joseph F.
;    APPLICANT: Lord, Reginald V.
;    APPLICANT: Alvarez, Chris
;    APPLICANT: Wetzel, Jon C.
;    APPLICANT: Scherf, Uwe
;    APPLICANT: Vockley, Joseph G.
;    TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
;    FILE REFERENCE: 44921-5026
;    CURRENT APPLICATION NUMBER: US/10/240,425
;    CURRENT FILING DATE:  2002-09-30
;    PRIOR APPLICATION NUMBER: PCT/US01/09847
;    PRIOR FILING DATE: 2001-03-28
;    PRIOR APPLICATION NUMBER: US 60/193,446
;    PRIOR FILING DATE: 2000-03-31
;    NUMBER OF SEQ ID NOS: 1588
;    SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1303
;    LENGTH: 1134
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    OTHER INFORMATION: Genbank Accession No. US20040033502A1 M20469
US-10-240-425-1303

Query Match             72.5%;  Score 387.8;  DB 7;  Length 1134;
Best Local Similarity   82.0%;  Pred. No. 5.9e-102;
Matches  483;  Conservative  0;  Mismatches  52;  Indels  54;  Gaps  1;
```

```
Qy      1  CGGCCCGCCTTCCTGGCCCAGCAGGAGAGCGAGATTGCTGGCATCGAATGACTCGGGTT  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243  CGGCCCGCCTTCCTGGCCCAGCAGGAGAGCGAGATTGCTGGCATCGAATGACTCGGGGCT  302

Qy     61  TCGGGGACACCTGCCGCGGCCAGGTGCCTCTGCGAGCGCGGGGTG  120
           ||||||||||||||||||||||||||||||||||||| |||||||
Db    303  TCGGGGACACCTGCCGCGGCCAGGTGCCATGCGGCAGCCGCAGCGAGTGGGGCTG  362

Qy    121  GTTCGGAGGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    363  GTTCTGAGGACATGGGGACCACCAGTCAATGGAGATGTGTTTCAGGAGGCCACGGGTCCTG  422

Qy    181  CCGATGGCTACGGCTGCCTGCCCAGGCCCAGGCGGACAGTTGACTCAGAGCATCC  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    423  CTGATGGCTACGGCAGCCATTGCCCAGGCTGACCCAGGCTGCAGAGCATCC  482

Qy    241  GCAAGTGCGAGAGGAGCAGAGAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGTGA  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    248  GAGAGAGGAGCAGAGAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGTGACCGAACA  307
```

```
RESULT 11
US-09-823-245A-213/c
; Sequence 213, Application US/09823245A
; Publication No. US2002003976OA1
; GENERAL INFORMATION:
;    APPLICANT: Wong, Gordon G.
;    APPLICANT: Clark, Hilary
;    APPLICANT: Fechtel, Kim
;    APPLICANT: Agostino, Michael J.
;    APPLICANT: Howes, Steven H.
;    APPLICANT: Resnick, Richard J.
;    APPLICANT: Gulukota, Kamalakar
;    APPLICANT: Graham, James R.
;    APPLICANT: Genetics Institute, Inc.
;    TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
;    FILE REFERENCE: GIN 6401
;    CURRENT APPLICATION NUMBER: US/09/823,245A
;    CURRENT FILING DATE:  2001-03-29
;    PRIOR APPLICATION NUMBER: 60/194,941
;    PRIOR FILING DATE: 2000-04-06
;    NUMBER OF SEQ ID NOS: 631
;    SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 213
;    LENGTH: 894
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-09-823-245A-213

Query Match             71.2%;  Score 380.8;  DB 3;  Length 894;
Best Local Similarity   81.8%;  Pred. No. 6e-100;
Matches  476;  Conservative  0;  Mismatches  52;  Indels  54;  Gaps  1;
```

```
Qy      8  CTTCCTGGCCCAGCAGGAGAGCGAGATTGCTGGCATCGAATGACTCGGGTTTCGGGC  67
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    894  CTTCCTGGCCCAGCAGGAGAGCGAGATTGCAGGCATAGAGAACGACGAGGGGCTTCGGGGC  835

Qy     68  ACCTGCCGCGGCCAGGTGCCTCTGCGAGCGCGGGGTG  127
           |||||||||||||||||||||||||||||||||||||
Db    834  ACCTGCCGCGGCCAGGTGCCATGCGGCAGCCGCAGCGAGTGGGGCTGGTTCTGA  775

Qy    128  GGACATGGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG  187
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    774  GGACATGGGGACCACCAGTCAATGGAGATGTGTTTCAGGAGGCCACGGGTCCTG  715

Qy    188  CTACGGCTGCCTGCCCAGGCCCAGGCGGACAGTTGACTCAGAGCATCC  247
           |||||||||||||||||||||||||||||||||||||||||||||||
Db    714  CTACGCAGCCATTGCCCAGGCTGACCCAGGCTGCAGAGCATCC  655

Qy    248  GAGAGAGGAGCAGAGAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGTGACCGAACA  307
```

Db   667 GCGAGAGGAGCAGAGGAAACGGCTGCAAGAGCTGGATCTGCGATCTCAAGGTCACGGAACA 608
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   308 GGAGTGGCGGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAGTGAACA 367
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   607 GGAATGGCGGGGAGGAGGCCAAGAGGACCTGGAGGAGTGGAACCAGCGCCAGAGTGAACA 548

Qy   368 GGTTGAGAAGAACAAGATCAACACAG------------------------ 394
          |||||||||||||||||||||||||||
Db   547 AGTAGAGAAGAACAAGATCAACACCGATCGCTGACAAAGCATTCTACCAGCAGCCAGA 488

Qy   395 ------------------GGCATCCGAAGAGAGGCTTTTGTGAAAGAATCCAAGGAGGA 433
                            |||||||||||||||||||||||||||||||||||
Db   487 TGCTGATATCATCGGCTACGTGGCATCCGAAGAGGCTTTCGTGAACGAAGGAATCCAAGGAGGA 428

Qy   434 GACCCCAGGCACAGAGTGGGAGAAGGTGCCCAGCTGTGTGACTTCAACCCTAAGAGCAG 493
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   427 GACCCCAGGCACAGAGTGGGAGAAGGTGCCCAGCTATGTGACTTCAACCCCAAGAGCAG 368

Qy   494 CAAGCAATGTAAAGACGTGTCCCGCCTGCGGTGCTCAT 535
          ||||||||||||||||||||||||||||||||||||||
Db   367 CAAGCAGTGCAAAGATGTGTCCCGCCTGCGGTCGGTGCTCAT 326

RESULT 13
US-09-918-995-20102
; Sequence 20102, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 20102
; LENGTH: 490
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(490)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-20102

Query Match   68.1%;   Score 364.4; DB 3; Length 490;
Best Local Similarity 89.3%;   Pred. No. 3e-95;
Matches 392; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy   97 AGCCCGGACTCGCCGAGCGGGGGTGGTTCGGAGGACATGGGGACTACAGTCAGTGAGATG 156
          |||| |||| |||| |||||||||||||||||||||||||||||||||||||||||||||
Db   44 AGCCGGGCCCCACGAGTGGGGCTGTTCTGAGGACATGGGGACCCACAGTCAGTGAGATG 103

Qy   157 TGTTTCAGGAGGCTAACGGCCTGAGATGCTACGGTGGGATTGCCAGGCGGACCAGGT 216
          |||||||||||||| ||||||||||||||||||| |||||||||||||||||||||
Db   104 TGTTTCAGGAGGCCAACGGTCCTGAGATGCTACGGTGGGATTGCCAGGCGGACAGGC 163

Qy   217 TGACTCAGGAGCCTGAGAGCATCCGCAAGTGAGAGTGGAGAGGAGGCTGCAGG 276
          |||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   164 TGACCCAGGAGCCTGAGAGCATCCGCAAGTGTGCGAGAGGAGGCGAAACGGGTGCAAG 223

Qy   277 AGTTGGATGCTGCCTCGAAGTGACCGAACAGAGGCCAAAAAGACC 336
          |||||||||||| |||||||||||||||||||||||||||||||||
Db   224 AGCTGGATGCTGCATCTAAGGTACGGACAACAAGAATGGCGGGGACCAGCAAGAAGGACC 283

Qy   337 TGGAGGAGTGGAACCAGGCCAAAGTGAACATGGCCCAAAGTG 396
          |||||||||||||||||| |||||||||||||||||||||||
Db   284 TGGAGGAGTGGAACCAGCGCCAGAGTGAACAAGTAGAGAAGAGATCAACAACAGG 343

Db   654 GCGAGAGGAGCAGAGGAAACGGCTGCAAGAGCTGGATCTGCGATCTCAAGGTCACGGAACA 595
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   308 GGAGTGGCGGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAGTGAACA 367
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   594 GGAATGGCGGGGAGGAGGCCAAGAGGACCTGGAGGAGTGGAACCAGCGCCAGAGTGAACA 535

Qy   368 GGTTGAGAAGAACAAGATCAACACAG--------------- 394
          ||||||||||||||||||||||||||
Db   534 AGTAGAGAAGAACAAGATCAACACCGATCGCTGACAAAGCATTCTACCAGCAGCCAGA 475

Qy   395 ------------------GGCATCCGAAGAGAGGCTTTTGTGAAAGAATCCAAGGAGGA 433
                            |||||||||||||||||||||||||||||||||||
Db   474 TGCTGATATCATCGGCTACGTGGCATCCGAAGAGGCTTTCGTGAACGAAGGAATCCAAGGAGGA 415

Qy   434 GACCCCAGGCACAGAGTGGGAGAAGGTGCCCAGCTGTGTGACTTCAACCCTAAGAGCAG 493
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   414 GACCCCAGGCACAGAGTGGGAGAAGGTGCCCAGCTATGTGACTTGACTTCAACCCCAAGAGCAG 355

Qy   494 CAAGCAATGTAAAGACGTGTCCCGCCTGCGGTGCTCAT 535
          ||||||||||||||||||||||||||||||||||||||
Db   354 CAAGCAGTGCAAAGATGTGTCCCGCCTGCGGTCGGTGCTCAT 313

RESULT 12
US-09-823-245A-178/c
; Sequence 178, Application US/09823245A
; Publication No. US20020039760A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6401
; CURRENT APPLICATION NUMBER: US/09/823,245A
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/194,941
; PRIOR FILING DATE: 2000-04-06
; SOFTWARE: PatentIn Ver. 2.0
; NUMBER OF SEQ ID NOS: 631
; SEQ ID NO 178
; LENGTH: 907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-823-245A-178

Query Match   70.9%;   Score 379.2; DB 3; Length 907;
Best Local Similarity 81.6%;   Pred. No. 1.7e-99;
Matches 475; Conservative 0; Mismatches 53; Indels 54; Gaps 1;

Qy   8 CTTCCTGGCCCAGCAGGAGAGCGAGATTGCTGGCATCGAGAATGACTCCGGGTTTCGGGGC 67
          |||||||||| ||||||||||||| ||| ||||||||||||||||||||||||||||||
Db   907 CTTCCTGGCCCAGCAGGAGAGCGAGATTGCAGGCGGCGAGACGACGAGGGGCTTCGGGGGC 848

Qy   68 ACCTGCCGCGACCAGGCCAGCCTGCCTCTGCGCAGCCGGGTGGTTCGGA 127
          |||||||||| ||||||||||| || ||||||||||||||| |||||||||
Db   847 ATCTGCCGGCAGCCAGCCTGCCCACGCGGCCCCGCAGCGAGTGGGGTCTGGTTCTGA 788

Qy   128 GGACATGGGGACTACAGTCAATGAGATGTGTTTCAGGAGGATCAACGGCCTGCCGATGG 187
          ||||||||| ||||||||||||||||||||||||||||| ||||||||||||||| |||||
Db   787 GGACATGGGGACCACAGTCAGTGAGATGTGTTTCAGGAGGCCAACGGTCCTGCTGATGG 728

Qy   188 CTACGCTGCGATTGGCACCGGCGGACAGGTTTGACTCAGGAGCCTGAGAGCATCCGCAAGTG 247
          |||||||||||||||||||||| |||||||| ||||||||||||||||||||||||||||||
Db   727 CTACGGCCAGCCATTGCCCAGGCTGACAGGCGCATCCCAGCTGACTCCGCCAAGTG 668

Qy   248 GAGAGGAGGAGCAGAGAAAAGGCTGCAAGAGCTGGATGCTGCCTCGAAGGTGACCGAACA 307

Qy   397 CATCGGAAGAGGCTTTTGTGAAAGAATCAAGGAGGAGACCCCAGGCACAGAGTGGGAGA 456
           ||||| |||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   344 CATCCGGAGGAGGCTTTCGTGAGGAGGAATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGA 403

Qy   457 AGGTGGCCCAGCTGTGTGACTTCAACCCTAGAGCAGCAGCCAAGCAATGTAAAGACGTGTCCC 516
           |||||||||||||||| |||||| ||||||| ||||||||||||||||||||||||||||||
Db   404 AGGTGGCCCAGCTATGTGACTTGACTTCAACCCCCAAGAGCAGCAGCGANGCAGTGCATAGATGTGTCCC 463

Qy   517 GCCTGCGCTCGGTGCTCAT 535
           |||||||||||||||||||
Db   464 GCCTGCGCTCGGTGCTCAT 482

RESULT 14
US-09-918-995-26855
; Sequence 26855, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;   APPLICANT: Hyseq, Inc.
;   TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;   TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
;   FILE REFERENCE: 20411-756
;   CURRENT APPLICATION NUMBER: US/09/918,995
;   CURRENT FILING DATE: 2001-07-30
;   PRIOR APPLICATION NUMBER: US/09/235,076
;   PRIOR FILING DATE: 1999-01-20
;   NUMBER OF SEQ ID NOS: 38054
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 26855
;   LENGTH: 493
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(493)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-26855

Query Match   54.9%;   Score 293.8; DB 3;   Length 493;
Best Local Similarity   80.7%;   Pred. No. 8.3e-75;
Matches 343; Conservative   0; Mismatches 82; Indels   0; Gaps   0;

Qy   39 GGGCATCGGAAGAGGCTTTTCGGGTTTCGGGCACCTGCCCGCCAGCTGGCCTCTGCGCCAG 98
           |||||||||||||||||||||||||||||||||||||||| |||||| |||||||||||||
Db   61 GGCATAGAGAACGACGAGGGCTTCGGGGCCCCCCGGCCCCAGCCCCGCCATGCCGCCGCCAG 120

Qy   99 CCCGGACTCGCCGAGCGGGGGTGGTTCGGAGGACATGGGACTACAGTCAATGGAGATGTG 158
           ||||||||||||||||||||| ||||||||||||||||||| ||||||||||||||||||
Db   121 CCGGGGCCCCACGAGTGGGGCTGGGGGCTGGTGTTCTGAGGACATGGGACCACAGTCAATGGAGATGTG 180

Qy   159 TTTCAGGAGGCTAACGGCCTGCCGATGGCTACGGCTGCCCAGGCGGACAGGTTG 218
           |||||||||||||||||||||| ||||||||||| |||||||||||||| |||
Db   181 TTTCAGGAGGCCCAACGTCCTGCGTGATGGCTACGCGCACCGAGCCCAGGCTGACAGGCTG 240

Qy   219 ACTCAGGAGCCTGAGAGCATCCGCAAGTGAGGAACAGCAGAGGGCTGCAGGAG 278
           |||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   241 ACCCAGGAGGAGGCCGTGAGAGCTGCCCAAGTGAGGCCAGAGAGGAGAAACGGCTGCAAGAG 300

Qy   279 TTTGGGATGCTGCCTCGAAGGTGACCGAACACAGAACAGTTGAGAAGAACAGACCTG 338
           |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   301 CTGGGATGCTGCATCTCAAGGTCACGGAACATGGCGGGATGGCCCGGGAGGAACAGGGACTG 360

Qy   339 GAGGAGTGGAACCAGCGCCAAAGTGAACAGTTGAGAAGAACAGACCAGGGCA 398
           |||||||||||||||||||||||||||||||||| ||||||| |||||||||
Db   361 GAGGAGTGGAACCAGCGCCAAAGTGAACAAGTTGAGAAGAGAGAGAGAGAACAAGATCAACAACACCGGATC 420

Qy   399 TCGGAAGAGGCTTTTGTGAAAGAATCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAG 458
           || |||||||| |||||||| |||||| ||||||||| |||||||||||||||||
Db   421 GCTGACAAAGCATTCTACCAGCAGCCAGATGCTGATATCATCCGGCTACGTGCGCATCCGAGG 480

Qy   459 GTGGC 463
           |||||
Db   481 GAGGC 485

RESULT 15
US-09-918-995-27337
; Sequence 27337, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;   APPLICANT: Hyseq, Inc.
;   TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;   TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
;   FILE REFERENCE: 20411-756
;   CURRENT APPLICATION NUMBER: US/09/918,995
;   CURRENT FILING DATE: 2001-07-30
;   PRIOR APPLICATION NUMBER: US/09/235,076
;   PRIOR FILING DATE: 1999-01-20
;   NUMBER OF SEQ ID NOS: 38054
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 27337
;   LENGTH: 485
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(485)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-27337

Query Match   50.2%;   Score 268.8; DB 3;   Length 485;
Best Local Similarity   77.1%;   Pred. No. 1.4e-67;
Matches 327; Conservative   0; Mismatches 97; Indels   0; Gaps   0;

Qy   40 GCATCGGAAATGACTCGGGTTTCGGGCACCTGCCCGCCAGCTGGCCTCTGCGCAGC 99
           |||||||| |||| |||||||||||||||||||||||||||||||||||||||||
Db   58 GCATACAGAACGACAAGGGCTTCTGGGCACCTGCCCGTCAGCGCATGCCATTCCCGCGCATT 117

Qy   100 CCGGGACTTCGCCGAGCGGGGTGGTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGT 159
           ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   118 CCGGGCCCCACGAGTGGGGCTGGGGGCTGGGGTGAGGACATGGGGACCACAGTCAATGGAGATGTGT 177

Qy   160 TTCAGGAGGCTAACGGCCTGCCGATGGCTACGGCTGCCGATTGCCCAGGCGGACAGGTTGA 219
           |||||||||||||||||||||| |||||||||| ||||||||||||||| ||||||| ||
Db   178 TTCAGGAGGCCCAACGTCCTGCGTGATGGCTACGCGCACCGAGCCCAGGCTGACAGGCTGA 237

Qy   220 CTCAGGAGCCTGAGAGCATCCGCAAGTGAGGAGCAGAGAAAAGGCTGCAGGAGT 279
           ||||||||||||||||| ||||||||||||||| ||||||||||||||||||||
Db   238 CCCATGAGGCTGAGAGCATTCGCAAGTGCAGTGCGGAGGAGGAGAATGCGCTGCAAGAGC 297

Qy   280 TGGATGCTGCCTCGAAGTGACCGAACAGGACCGAAAAAAAGACCTGG 339
           |||||||||||||||||||||||||||||| |||||||||||||| ||
Db   298 TGGATGCTGCATCTAAGGTCACGGAGAAGCAGGAGGCAAGAATGAACTGACCTGG 357

Qy   340 AGGAGTGGAACCAGCGCCAAAGTGAACAGTTGAGAAGAACAACAGAGGGCAT 399
           |||||||||||||||||||| |||||| |||||||||| |||||||||||||
Db   358 AGGAGTGGAACCAGCGCTGCTAGTGAACAAGTTTAGAAGAACAAGATCAACAACCGGATCG 417

Qy   400 CGGAAGAGGCTTTTGTGAAAGAATCCACCAGCAGCCCAGGCGTGGGAGAAGG 459
           |||||||| |||||||| |||| |||||||||||||||||| ||||||||||
Db   418 CTGCACAAGCATTCTACCAGCAGCCAGATGCTGATATCATCGGCTACGTGCGCATCCGAGG 477

Qy   460 TGGC 463
           ||||
Db   478 AGGC 481

Search completed: January 5, 2006, 21:49:02
Job time : 807 secs

```
              GenCore version 5.1.6
       Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    January 5, 2006, 19:41:26 ; Search time 312 Seconds
                             (without alignments)
                             1249.941 Million cell updates/sec

Title:     US-10-664-705-145
Perfect score: 535
Sequence:  1 cggccgccttcctggcccag.........cgcctgcgctcggtgctcat 535

Scoring table: IDENTITY_NUC
           Gapop 10.0 , Gapext 1.0

Searched:  4637609 seqs, 364468668 residues

Total number of hits satisfying chosen parameters:    9275218

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Published Applications_NA_New:*
        1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
        2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
        3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
        4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
        5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
        6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
        7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
        8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
        9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
       10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Match Length	DB	ID	Description
1		514.6	96.2	946	7	US-11-136-527-1696	Sequence 1696, Ap
2	c	481.6	90.0	600	7	US-11-136-527-5792	Sequence 5792, Ap
3		213.4	39.9	657	6	US-10-821-234-589	Sequence 589, App
4		213.4	39.9	1023	6	US-10-955-054A-94	Sequence 94, Appl
5		146	27.3	1169	7	US-11-136-527-347	Sequence 347, App
6		146	27.3	1169	7	US-11-136-527-4443	Sequence 4443, Ap
7		89.6	16.7	2737	6	US-10-955-054A-123	Sequence 123, App
8		57	10.7	153376	7	US-11-121-086-5	Sequence 5, Appli
9	c	49.2	9.2	120096	7	US-11-121-086-24	Sequence 24, Appl
10		47.8	8.9	207908	7	US-11-112-908-21	Sequence 21, Appl
11		47.8	8.9	212805	7	US-11-112-908-19	Sequence 19, Appl
12		45.4	8.5	168516	7	US-11-121-086-3	Sequence 3, Appli
13	c	45.2	8.4	1400	7	US-11-136-527-7472	Sequence 7472, Ap
14		45.2	8.4	2380	7	US-11-136-527-3376	Sequence 3376, Ap
15	c	44	8.2	5468	6	US-10-821-234-49	Sequence 49, Appl
16		42.8	8.0	2099	7	US-11-136-527-2825	Sequence 2825, Ap
17		42.4	7.9	173602	7	US-11-121-086-25	Sequence 25, Appl
18		41.6	7.8	8280	7	US-11-000-688-641	Sequence 641, App
19		40.4	7.6	1400	7	US-11-136-527-4440	Sequence 4440, App
20		40.4	7.6	2305	7	US-11-136-527-344	Sequence 344, App
21		39.4	7.4	26667	6	US-10-995-561-13375	Sequence 13375, A
22		39.2	7.3	1027	7	US-11-136-527-94	Sequence 94, Appl
23		39.2	7.3	2752	7	US-11-136-527-2852	Sequence 2852, Ap
24	c	39.2	7.3	28037	6	US-10-995-561-13220	Sequence 13220, A
25	c	38	7.1	1619	7	US-10-821-234-10	Sequence 10, Appl
26		37.8	7.1	3897	7	US-11-136-527-2555	Sequence 2555, Ap
27		37.6	7.0	1400	7	US-11-136-527-4371	Sequence 4371, Ap
28		37.6	7.0	5941	7	US-11-136-527-275	Sequence 275, APP
29	c	37.2	7.0	1991	7	US-11-136-527-3217	Sequence 3217, Ap
30	c	37	6.9	201	6	US-10-995-561-21286	Sequence 21286, A
31		37	6.9	4457	6	US-10-775-169-259	Sequence 259, APP
32		37	6.9	6025	6	US-10-995-561-143	Sequence 143, APP
33		37	6.9	6148	6	US-10-995-561-147	Sequence 147, APP
34		37	6.9	6185	6	US-10-995-561-146	Sequence 146, APP
35		37	6.9	6187	6	US-10-995-561-144	Sequence 144, APP
36		37	6.9	6886	6	US-10-995-561-149	Sequence 149, APP
37		37	6.9	6925	6	US-10-995-561-145	Sequence 145, APP
38		37	6.9	171486	7	US-11-121-086-105	Sequence 105, APP
39		37	6.9	179777	7	US-11-121-086-106	Sequence 106, APP
40		36.4	6.8	91561	7	US-11-124-368A-2896	Sequence 2896, Ap
41		36.2	6.8	6900	7	US-11-000-688-1142	Sequence 1142, Ap
42		36.2	6.8	8948	6	US-10-623-155-119	Sequence 119, APP
43	c	36.2	6.8	11438	7	US-10-821-234-224	Sequence 224, APP
44		36.2	6.8	11612	6	US-10-995-561-499	Sequence 499, APP
45		36.2	6.8	11684	6	US-10-995-561-498	Sequence 498, APP

ALIGNMENTS

```
RESULT 1
US-11-136-527-1696
;  Sequence 1696, Application US/1136527
;  Publication No. US20050287570A1
;  GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT: Mounts, William M
;  TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
;  FILE REFERENCE: 031896-041000 (AM101086)
;  CURRENT APPLICATION NUMBER: US/11/136,527
;  CURRENT FILING DATE: 2005-05-25
;  PRIOR APPLICATION NUMBER: US 60/574,294
;  PRIOR FILING DATE: 2005-05-26
;  NUMBER OF SEQ ID NOS: 362830
;  SOFTWARE: PatentIn version 3.2
;  SEQ ID NO 1696
;  LENGTH: 946
;  TYPE: DNA
;  ORGANISM: Rattus norvegicus
US-11-136-527-1696

  Query Match             96.2%;  Score 514.6;  DB 7;  Length 946;
  Best Local Similarity   90.5%;  Pred. No. 1.4e-126;
  Matches 484;  Conservative 51;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1  CGGCCGCCTTCCTGGCCCAGGAGGAGAGCGAGATTGCTGCATCGAGAATGACTCGGGTT   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     99  CGGCCGCCTTCCTGGCCCAGGAGGAGAGCGAGATTGCTGCATCGAGAATGACTCGGGTT  158

Qy     61  TCGGGGCACCTGCGCCTCGCCAGGTGGCCTCTGCGCCGCCAGGTGGCCTCTGCGCGGGGGTG  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    159  TCGGGGCACCTGCGCCTCGCCAGGTGGCCTCTGCGCCGCCAGGTGGCCTCTGCGCGGGGGTG  218

Qy    121  GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGTCAACGGCCTG  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219  GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGTCAACGGCCTG  278

Qy    181  CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCTCAGGAGCATCC  240
           |||||||||||||||||||||||||||||||||:|::|::| |||| |||||||||||||
Db    279  CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGASKCASGAGSYTGAGAGCATCC  338

Qy    241  GCAAGTGGAGGAGAGAGCAGAAGAAAAGGCTGCAGGAGAGTGCTGCCTCGAAGTGA  300
           ||||||||||||||||||:|::|::||::| :::|:: ||||:||||||||:::::
Db    339  GCAAGTGGAGGAGAGAGCGRGARGAGAGCGCRGARGAGAGCGCWKRAGTTGGATGCTGCCTCGAWGGYKA  398

Qy    301  CCGAACAGGAGTGGGCGGGAGAAGGCCAAAAAGACCTGGAACGACGCCAAA  360
```

```
Db   399 CCGAACWGGAKTGSCSRGAGMAGSMCARRWWAGACCWGGAGSMGTGRRASCAKCSCCAAA 458
         ||||||||:||||||:||::||::||||:|||||
Qy   361 GTGAACAGGTTGAGAAGAACAAGATCAACAACAGAGTGGCCCAGCTTTTGTGAAAG 420
         ||||||:||||::||:::||
Db   459 GTGAACAGGTWRASMWGWMCWMKMTCAMCWWCAGGGCATCGGAAGAGGCTTTKTGAAAG 518
Qy   421 AATCCAAGGAGGAGACCCCAGCCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCA 480
         ||||||:||||||:|||:|||||||||||||||||||||||||||||||||||||
Db   519 AATCCAAGGWGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACYTCA 578
Qy   481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTGCGCCTCGGTGCTCAT 535
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   579 ACCSTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTGCGCCTCGGTGCTCAT 633
```

RESULT 2
US-11-136-527-5792/c
; Sequence 5792, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5792
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-5792

Query Match   90.0%; Score 481.6; DB 7; Length 600;
Best Local Similarity   89.8%; Pred. No. 6.5e-118;
Matches 451; Conservative 51; Mismatches 0; Indels 0; Gaps 0;

```
Qy     1 CGGCCGCCTTCCTGGCCCAGCAGGAGAGCCAGATTGCTGGCATCGAGAATGACTCGGGTT 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   502 CGGCCGCCTTCCTGGCCCAGCAGGAGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT 443
Qy    61 TCGGGGCACCTGCCGCCCAGGTGGCCTCTGCGCAGCCCCGACTCGCGAGGGGGGGTG 120
         ||||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db   442 TCGGGGCACCTGCCGCCCAGGTGGCCTCTGCGCAGCCCCGACTCGCGAGCGGGGGTG 383
Qy   121 GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   382 GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG 323
Qy   181 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGAGCATCC 240
         |||||||||||||||||||||||||||:||||:|||||||||:|||||||
Db   322 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGASKCASGAGSYTGAGAGCATCC 263
Qy   241 GCAAGTGAGAGACATGGGAGTGAGAGAAGAAAGGCTCAGGAGGCTAACGGGCCTGA 300
         ||||:||||||||:|:||||||||||||||||||||||||:||:||
Db   262 GCAAGTGAGAGAGTGCGGAGAAGCGRGARGAAAAGGCTGCWKRAGTTGGATGCTGCCTCGAWGGYKA 203
Qy   301 CCGAACAGGAGTGCGGAGAAGGCCAAAAAGACCTGGAGAGTGAACCAGCGCCAAA 360
         |||:|:|||||:||:|||::||
Db   202 CCGAACWGGAKTGSCSRGAGMAGSMCARRWWAGACCWGGAGSMGTGRRASCAKCSCCAAA 143
Qy   361 GTGAACAGGTTGAGAAGAACAACAGAGTGGCCCAGCTTTTGTGAAAG 420
         ||||||:|||||:||:::||||:
Db   142 GTGAACAGGTWRASMWGWMCWMKMTCAMCWWCAGGGCATCGGAAGAGGCTTTKTGAAAG 83
Qy   421 AATCCAAGGAGGAGACCCCAGCCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCA 480
         |||||||:||||:|||::||||||||||||||||||||||||||:|||||||||:|
Db    82 AATCCAAGGWGAGACCCCAGGCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACYTCA 23
```
```
Qy   481 ACCCTAAGAGCAGCAAGCAATG 502
         |||||||||||||||||||:|||
Db    22 ACCSTAAGAGCAGCAAGCAATG 1
```

RESULT 3
US-10-821-234-589
; Sequence 589, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 589
; LENGTH: 657
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-821-234-589

Query Match   39.9%; Score 213.4; DB 6; Length 657;
Best Local Similarity   63.7%; Pred. No. 3.6e-47;
Matches 341; Conservative 0; Mismatches 191; Indels 3; Gaps 1;

```
Qy     1 CGGCCGCCTTCCTGCGCCAGCAGGAGAGCCAGATTGCTGGCATCGAGAATGACTCGGGTT 60
         ||||||||||||||:|||||||||||||:||||||:||||||||:||||:|||||||||
Db    95 CTGCGGCCCTTCTTGGCGGACGGGCGGC---GCCCCCGGGCCCAGCCCCAGCGGCCGCCGGGGCCGCCTAACGGCCCTG 154
Qy    61 TCGGGGCACCTGCGCGCCCAGGTGGCCTCTGCGCAGCCCCGGGCGCAGCCGCGGGGGTG 120
         ||||:||||:||||||:|:||||||||||||||||||:||||||||||:||||||||
Db   155 TCGGGGCACTCCTGGACGGGCGGC---GCCCCCGGGCCCAGCCCCAGCCGGAAAGTAATAGTCCAA 211
Qy   121 GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG 180
         |||||||:|||||||||||:|:|||||||:||||:|||:|||||||:|||||:||||
Db   212 GTTCGGAGGACATGGGACTACAGTGAATGAGTAATGGTGAATACTACCAGGAAAGTAATGGTCCAA 271
Qy   181 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGAGCATCC 240
         ||||||||||:||:|||||||||||||||||||||:|||||:|||||||||||||
Db   272 CAGACAGTTATGCAGCTCATTTCACAAGTGGATCGATGCAGTCAGAGCCTGAAAGTATCC 331
Qy   241 GCAAGTGAGAGAGGAGCAGAGAAGAAAGGCTCGAGGAGTTGGATGCTGCCTCGAAGGTGA 300
         ||||||||||||:||:|||:||||||||:||||:|||||:|||:||||:||||:|||
Db   332 GTAAATGGAGAAGAAGAACAAATGAACGCTTGGAAGCCCCTTGATGCCAATTCTCGGAAGC 391
Qy   301 CCGAACAGGAGTGCGGGAGAAGGCCAAAAAGACAAGATCAACAACAGGGCATCGGAAGATCC 360
         |||||||||||:||:|||||:|||:|||:|:|||||||||||:|||||||:|||||||||
Db   392 AAGAAGCAGAGTGGAAGAAAAGGCAATAAAGGCTAGAAGAATGGTATGCAAGACAGG 451
Qy   361 GTGAACAGGTTGAGAAGAACAACAAGATCAACAACAGAGTGGCCCAGCTTTTGTGAAAG 420
         ||||||||||:||:|||:||:|||||||:|:|||:|:|||||:|||||||::|||:||
Db   452 ACGAGCAGCTACAGACAAAAGCAACAAAAGCAACAACAGGGCCAGCAGAAGAAGCCTTTGTAAAATG 511
Qy   421 AATCCAAGGAGGAGACCCCAGCCACAGAGTGGGAGAAGGTGGCCCAGCTGTGTGACTTCA 480
         ||:|::|:||:||||:||||:||:||:|||||:|:|||:||:|||:|||||:||||||
Db   512 ACATTGACGAGTCGTCCCAGGCACTGAGTGGGACGGGTGGCCCCGGCGTCGGTGTGACTTTA 571
Qy   481 ACCCTAAGAGCAGCAAGCAATGTAAAGACGTGTCCCGCCTCGGTGCTCAT 535
         |||||||||||||||||||||||||:||||||||||||||:||:|:||
Db   572 ACCCCAAGTTCTAGCAGCAGGCCAAAGATGTCTCCCGCCTCAGTCGTTCAGTCTCAT 626
```

RESULT 4
US-10-955-054A-94
; Sequence 94, Application US/10955054A
; Publication No. US2005026420A1

; GENERAL INFORMATION:
; APPLICANT: PUSZTAI, LAJOS
; APPLICANT: SYMMANS, W. FRASER
; APPLICANT: HESS, KENNETH R.
; APPLICANT: AYERS, MARK
; APPLICANT: STEC, JAMES
; TITLE OF INVENTION: MULTIGENE PREDICTORS OF RESPONSE TO CHEMOTHERAPY
; FILE REFERENCE: UTKC:880US
; CURRENT APPLICATION NUMBER: US/10/955,054A
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 195
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 94
; LENGTH: 1023
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-955-054A-94

Query Match   39.9%;   Score 213.4; DB 6; Length 1023;
Best Local Similarity 63.7%;   Pred. No. 4e-47;
Matches 341; Conservative   0; Mismatches 191; Indels   3; Gaps   1;

Qy   1 CGGCCGCCTTCCTGCCCAGCAGGAGGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT   60
           ||||| || |||| |||||||||| ||||| |||||| ||||||||| ||||||||||||
Db   177 CTGCGGGGCCCTTCTTGGCGCAGCAAGAGAGCGAGATTGCGGGCATCGAGAGACGAGGCCT 236

Qy   61 TCGGGGCACCTGCCGCCAGCCTCTGCGCAGCCCGGACTCCGCAGCCCGAGCCGCCGGGTG 120
           ||| ||||| |||||| || ||||||||||||||||||| ||||
Db   237 TCGCCATCCTGGACGGCGGC---GCCCCGGGGCCCCAGCGCGACGGGCGAGCCGCCGGGGG 293

Qy   121 GTTCGGAGGACATGGGACTACAGTCAATGAGATGTGTTTCAGGAGGCTAACGGGCCTG 180
           |||| |||| ||||| ||||| |||||||||||| | || |||||||||| ||||
Db   294 GTCCGGATGCTGTTGATGAGTAATGATGTGAATACTACCAGGAAAGAAAGTAATGGTCCAA 353

Qy   181 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGGTTGACTCAGGAGCCTGAGAGCATCC 240
           ||||||||||| || |||| ||| ||||||||||| | ||| ||||||||||||||||
Db   354 CAGACAGTTATGCAGCTATTTCACAAGTGATCGATTGCGAGCCTGAAAGTATCC 413

Qy   241 GCAAGTGGAGAGAGGAGCAGAAGAAAGGCTGCAGGAGTTGATGCTGCCTCGAAGGTGA 300
           |||||||||||| |||||||||||||||| | |||||| || ||||||||||||
Db   414 GTAAATGAGAGAAGAGAAGCAAAATGAAGCCTTGGAAGCCCTTGAATGCCAATTCTCGGAAGC 473

Qy   301 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTAGAAGAATGGTATGCAAGACAGG 360
           ||| || | | | | || |
Db   474 AAGAAGCAGAGTTGAGAAGAACAAGAATCAACAACAGGGCATCGGAAGAGAGCTTTTGTGAAAG 533

Qy   361 GTGAACAGGTTGAGAGAAGAACAAGATCAACAACAGGGCATCGGAAGAGAGCTTTGTGAAAG 420
           || ||| ||||| |||||| ||||||||||
Db   534 ACGAGCAGCTACAGAAAACAAACAAACAAACACAGGGCAGCAGAAGAAGAGCCTTTGTAAATG 593

Qy   421 AATCCAAGGAGGAGACCCCAGGCACAGAGTGGGGACAAGGTGGCCCAGCTGTGTGACTTCA 480
           |||||||||| |||||||| |||||||| || |||||||||||| ||||
Db   594 ACATTGACGAGTCGTCCCCAGGCACTGAGTGGGAACGGGTGGCCCGGCTGTGTGACTTTA 653

Qy   481 ACCCTAAGACAGCAGCAATGTAAAGACGTGTCCCGCCTGGCTCGGTGCTCAT 535
           ||||||| | | | || | | | | |
Db   654 ACCCCAAGTCTAGCAAGCAGCAGCCCAAAGATGTCTCCCGCATGCGCTCAGTCCTCAT 708

RESULT 5
US-11-136-527-347
; Sequence 347, Application US/11136527
; Publication No. US2005028757OA1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830

; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 347
; LENGTH: 1169
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-347

Query Match   27.3%;   Score 146; DB 7; Length 1169;
Best Local Similarity 56.8%;   Pred. No. 2.5e-29;
Matches 259; Conservative   14; Mismatches 180; Indels   3; Gaps   1;

Qy   1 CGGCCGCCTTCCTGCCCAGCAGGAGGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT   60
           ||||| || |||| |||||||||| ||||| |||||| ||||||||| ||||||||||||
Db   210 CCGCGGGGCCTTCTTGGCGCAGCAAGAGAGTGAGATTGCGGGCATCGAGAGACGAGGCCT 269

Qy   61 TCGGGGCACCTGCCGCCAGCCTCTGCGCAGCCCGGACTCCGCAGCCCGAGCCGCCGGGTG 120
           ||| ||||| |||||| || |||||||||||||| || ||||
Db   270 TCGCCATCCTGACGGCGGC---GCCCCGGGGACACACGGCGAGCCGCCGCCGGGGG 326

Qy   121 GTTCGGAGGACATGGGACTACAGTCAATGAGATGTGTTCAGGAGGCTAACGGGCCTG 180
           |||| |||| |||| ||||| |||||||| | || |||||||||||||| ||
Db   327 GTCCTGATGCGGTTGATGGAGTGATGAATACTACCAGGAGAGAGGCCAATGGTCCAA 386

Qy   181 CCGATGGCTACGCTGCGATTGCCCAGGCGGACAGTTGACTCAGGAGCCTGAGAGCATCC 240
           ||||||||||| || |||| ||| | |||||||| || ||| ||||||||||||||||
Db   387 CAGACAGTTACGCAGCCATTTCAGAGTGATCGGTTGCAGTCAGCCTGAAAGTATCC 446

Qy   241 GCAAGTGGAGAGAGGAGCAGAAGAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGA 300
           |||||||||||| |||||||||||||||| | |||||| || ||||||| |||||||
Db   447 GTAAGTGGAGAAGAGCAGACAGGGCGCCTGAGAGCCCCTGGAGAGCCCTCGATGCCAATTCTCGAAGC 506

Qy   301 CCGAACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTTGAACCAGCGCCAAA 360
           ||| || ||||| |||| |||||| |||||||| ||||||||||||||||||||||
Db   507 AGGAGCAGAGTGGAAAGAAAGAAAGGCAGTGAAGAGCTGGAAGAGTTGTACGCAAGGCAGG 566

Qy   361 GTGAACAGGTTGAGAGAAGAACAAGATCAACACACAGGGCATCGGAAGAGAGCTTTTGTGAAAG 420
           || ||| ||||| |||||| ||||||||||
Db   567 ATGAGCAGCTACAGAAGAGAACAGCAGCAAAAGCCAGCAGATGAAGCTTTCYRYRAAS 626

Qy   421 AATCCAAGGAGGAGACCCCAGGCCACAGAGTGGGGAGA 456
           :|:::| ::::: :::::|
Db   627 RASCCTKCGMWGACCTGRTWSGYWADGKCASRAMSA 662

RESULT 6
US-11-136-527-4443
; Sequence 4443, Application US/11136527
; Publication No. US2005028757OA1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4443
; LENGTH: 1169
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-4443

Query Match   27.3%;   Score 146; DB 7; Length 1169;
Best Local Similarity 56.8%;   Pred. No. 2.5e-29;
Matches 259; Conservative   14; Mismatches 180; Indels   3; Gaps   1;

Qy   1 CGGCCGCCTTCCTGCCCAGCAGGAGGAGCGAGATTGCTGGCATCGAGAATGACTCGGGTT   60
           ||||| || |||| |||||||||| ||||| |||||| ||||||||| ||||||||||||
Db   210 CCGCGGGGCCTTCTTGGCGCAGCAAGAGAGTGAGATTGCGGGCATCGAGAGACGAGGCCT 269

```
;   APPLICANT: POULSEN, TIM S.
;   APPLICANT: NIELSEN, KIRSTEN V.
;   TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
;   FILE REFERENCE: 09138.6000-00000
;   CURRENT APPLICATION NUMBER: US/11/121,086
;   CURRENT FILING DATE: 2005-05-04
;   PRIOR APPLICATION NUMBER: 60/567,570
;   PRIOR FILING DATE: 2004-05-04
;   NUMBER OF SEQ ID NOS: 107
;   SOFTWARE: PatentIn version 3.3
;   SEQ ID NO 5
;     LENGTH: 153376
;     TYPE: DNA
;     ORGANISM: Homo sapiens
US-11-121-086-5

Query Match             10.7%;   Score 57;   DB 7;   Length 153376;
Best Local Similarity   48.9%;   Pred. No. 2.5e-05;
Matches  153;   Conservative   0;   Mismatches 160;   Indels    0;   Gaps    0;

Qy     163  AGGAGGCTAACGGCCTGCCGATGGCTACGCTCGGATTGCCCAGGCGACAGGTTGACTC  222
            |||||| |||||  |||||| | ||||||||||||||||||||||||||||||
Db   16026  AGGAGGGAGGAGGAACAGGAGGAGGAACAGGAGGAGGAACAGGAGGAGGAGGAGGAAC  16085

Qy     223  AGGAGGCCTGAGAGCATCCGCCAAGTGGACAGGAGGAGGACAGGAGAAAAGGCTGCAGGAGTTGG  282
            |||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   16086  AGGAGGAGGAGGAGGAGGAACAGGAGGAGGAGGAGGACAGGAGGAGGAGGACAGGAGGAGG  16145

Qy     283  ATGCTGCCTCGAAGGTGACCGAACAGGAGTGCCGGGAGGAGGAGGACAGGACAGGAGGCCAAAAAAGACCTGGAGG  342
            |||| || |||||||||||||||||||||||||||||||||||||||
Db   16146  AGGAACAGGACAGGAGGAGGCCACAGGAAGGCACAGGAGGAGGAGG  16205

Qy     343  AGTGGAACCAGCGCCAAAGTGAACAGGTTGAGAAGAACAAGATCAACACAGGGCATCGG  402
            |||||||||||||||| |||||||||||||||||
Db   16206  AACAGGAGGAGGAGGAACAGGAGGAGGACCCCAGGCACAGAGTTGGGAGAAGGTG  16265

Qy     403  AAGAGGCTTTTGTGAAAGAATCCAAGGAGGAGATCCAAGGAGGAGGAGGGTG  462
            |||||||   ||||||||
Db   16266  AGGAGGAGGAGGAGGAGCCACAGGAGGAGGAGGCAGCCACAGGAGGAGGAGG  16325

Qy     463  CCCAGCTGTGTGA  475
            ||||||||||
Db   16326  AACAGGAGGAGGA  16338
```

RESULT 9
US-11-121-086-24/c
;   Sequence 24, Application US/11121086
;   Publication No. US20050266459A1
;   GENERAL INFORMATION:
;   APPLICANT: POULSEN, TIM S.
;   APPLICANT: NIELSEN, KIRSTEN V.
;   TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
;   FILE REFERENCE: 09138.6000-00000
;   CURRENT APPLICATION NUMBER: US/11/121,086
;   CURRENT FILING DATE: 2005-05-04
;   PRIOR APPLICATION NUMBER: 60/567,570
;   PRIOR FILING DATE: 2004-05-04
;   NUMBER OF SEQ ID NOS: 107
;   SOFTWARE: PatentIn version 3.3
;   SEQ ID NO 24
;   LENGTH: 120096
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-121-086-24

Query Match   9.2%;   Score 49.2;   DB 7;   Length 120096;
Best Local Similarity   52.4%;   Pred. No. 0.0027;
Matches 108;   Conservative   0;   Mismatches 98;   Indels   0;   Gaps   0;

Qy   203 CCAGGCGACAGGTTGACTCAGGAGCCTGAGCGCTTCAGGAGCGATCCGCCAAGTGGAGAGAGGAGGAGCAGAA 262
            ||||| ||||| | |||| |||||||||||||||||||||||||||
Db   41948 CGAGGAGGAGGAGGAAGCCGAGAGAGGCGGAAAAAAGAGGAGGGGAAAAAAGAAACGGAGCA 41889

```
Qy      61  TCGGGGCACCTGCCGCCAGGTGGCTTCTGCGCAGCCCGACTCGCGAGCGGGGGTG  120
            |||||                           : ||||
Db     270  TCGCCCATCCTAGACGGCGGC---GCCCCCGGCGCCAGGMACACGGCGAGCCCGCCGGGGG  326

Qy     121  GTTCGGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGGCCTG  180
            ||||||||   ||||||  |  |||   ||  ||
Db     327  GTCCTGATGCGGTTGATGGAGTGATGAATACTACCAGGAGCAATGGTGTCCAA  386

Qy     181  CCGATGGCTACGCTGCCGATTGCCCAGGCGACAGTTGCCTCAGGAGGCCTGAGAGCATCC  240
            |||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db     387  CAGACAGTTACGCAGCCATTTCAGAAGTGGATCGGTTTCAGTCAGAGCCTGAAAGTATCC  446

Qy     241  GCAAGTGAGCAGGAGGAGGAGCAGAGAAAGGCTGCAGAGGTGTGCTGCCTCGAAGGTGA  300
            |                           ||||||||
Db     447  GTAAGTGGAGAAGAGCAGAGACAGAGCGCCTGAAGCCCTCGATGCCAATTCTCGGAAGC  506

Qy     301  CCGAACAGGAGTGCGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCAGCGCCAAA  360
            |||||||   ||||||||||||||||||||||||||
Db     507  AGGAAGCAGAGAGTGGAAAGAAAAAGGCAGTGAAGGAGGAGAGTGTACGCAAGGCAGG  566

Qy     361  GTGAACAGGTTGAGAAGAACAAGATCAACACAGGGCATCGGAAGAGGCTTTTGTGAAAG  420
            ||||||||||||||||||| || |||||||||||||||     :||:
Db     567  ATGAGCAGCTACAGAGACAAAAGCCAGCAACAGGGTCGGCAGATGAAGCTTTCYRYRAAS  626

Qy     421  AATCCAAGGAGGAGACCCCAGGCACAGAGTGGAGA  456
            :||::  ||  |:  ||
Db     627  RASCCTKCGMWGACVGTGRTWSGYWADGKCASRAMSA  662
```

RESULT 7
US-10-955-054A-123
;   Sequence 123, Application US/10955054A
;   Publication No. US20050266420A1
;   GENERAL INFORMATION:
;   APPLICANT: PUSZTAI, LAJOS
;   APPLICANT: SYMMANS, W. FRASER
;   APPLICANT: HESS, KENNETH R.
;   APPLICANT: AYERS, MARK
;   APPLICANT: STEC, JAMES
;   TITLE OF INVENTION: MULTIGENE PREDICTORS OF RESPONSE TO CHEMOTHERAPY
;   FILE REFERENCE: UTXC:880US
;   CURRENT APPLICATION NUMBER: US/10/955,054A
;   CURRENT FILING DATE: 2004-09-30
;   NUMBER OF SEQ ID NOS: 195
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 123
;   LENGTH: 2737
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-955-054A-123

Query Match   16.7%;   Score 89.6;   DB 6;   Length 2737;
Best Local Similarity   76.4%;   Pred. No. 2.3e-14;
Matches 110;   Conservative   0;   Mismatches 34;   Indels   0;   Gaps   0;

```
Qy     392  CAGGGCATCGGAAGAGGCTTTTGTGAAAGAATCCAAGGAGGAGACCCCAGGCACAGAGTG  451
            |||||||||||||   ||||||||||  ||  | |||||||||||||||||
Db    2317  CAGGGCAGCAGAAGAGCCTTTGTAAATGACATTGACCGATCGTCCCCAGGCACTGAGTG  2376

Qy     452  GGAGAAGGTGGCCCAGCTGTGTGACTTCAACCCTAAGAGCAGCAAGCAATGTAAAGACGT  511
            ||||||||||||||||||||||||||||||||||||||||||
Db    2377  GGAACGGGTGGCCCGGCCGGGGCCGCGTGTGACTTTTAACCCCAAGTCTAGCAAGCAGGCCAAAGATGT  2436

Qy     512  GTCCCGGCCTGCGCTCGGTGCTCAT  535
            |||||||||||||||||||||
Db    2437  CTCCCGCCATGCCGCTCAGTCCTCAT  2460
```

RESULT 8
US-11-121-086-5
;   Sequence 5, Application US/11121086
;   Publication No. US20050266459A1
;   GENERAL INFORMATION:

```
Qy    263  GAAAAGGCTGCGCAGGAGTTGGATGCTGCCTCGAAGGTGACCGACAGAGTGTGGCGGGAGAA  322
           |||  ||||| |||
Db  41888  GAAGGAGTTGGAGGAGGAGGTGGAGGAGGAGGAGGAAATGGAGCAGGAGGAAGAACAA  41829

Qy    323  GGCCAAAAAGACCTGGACGAGTGGAACCAGCGCCAAAGTGACCAGTTGAGAAGAACAA  382
           ||||| |||||| ||
Db  41828  GGAGGGAGAAGAAATGAAGAGGAGGAGCCAGCGCCAGAGAAATGAAGAGGAGGAGGA  41769

Qy    383  GATCAACAACAGGGCATCGGAAGAGG  408
           |||  || |||
Db  41768  GGAAAAAGAAATGAGGAGGAAAAAG  41743
```

RESULT 10
US-11-112-908-21
; Sequence 21, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 21
; LENGTH: 207908
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-21

Query Match   8.9%;   Score 47.8;   DB 7;   Length 207908;
Best Local Similarity   57.8%; Pred. No. 0.0072;
Matches   85;   Conservative   0;   Mismatches   62;   Indels   0;   Gaps   0;

```
Qy    249  AGAGGAGGAGCAGAGAAAAGGCTGCGCAGGAGTTGGATGCTGCCTCGAAGGTGACCGAACAG  308
           ||||||||| ||||||  | ||
Db 199013  AGAGGAGGAACAGAGAAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 199072

Qy    309  GAGTGCGGGAGAAGACCTGGACGAGTGGAACCAGCGCCAAAGTGAACAG  368
           |||| |||| |||| ||||
Db 199073  AAGGAGAAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGAG 199132

Qy    369  GTTGAGAAGAACAAGATCAACAACAGG  395
           |||||| ||| || |||||
Db 199133  GATAAGGAGAAGGAGAAGAACAAGAG 199159
```

RESULT 11
US-11-112-908-19
; Sequence 19, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 19
; LENGTH: 212805
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-19

Query Match   8.9%;   Score 47.8;   DB 7;   Length 212805;
Best Local Similarity   57.8%; Pred. No. 0.0072;
Matches   85;   Conservative   0;   Mismatches   62;   Indels   0;   Gaps   0;

```
Qy    249  AGAGGAGGAGCAGAGAAAAGGCTGCGCAGGAGTTGGATGCTGCCTCGAAGGTGACCGAACAG  308
           ||||||||| ||||||  | ||
Db  54934  AGAGGAGGAACAGAGAAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG  54993

Qy    309  GAGTGCGGGAGAAGACCTGGACGAGTGGAACCAGCGCCAAAGTGAACAG  368
           |||| |||| |||| ||||
Db  54994  AAGGAGAAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGAG  55053

Qy    369  GTTGAGAAGAACAAGATCAACAACAGG  395
           |||||| ||| || |||||
Db  55054  GATAAGGAGAAGGAGAAGAACAAGAG  55080
```

RESULT 12
US-11-121-086-3
; Sequence 3, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3
; LENGTH: 168516
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-3

Query Match   8.5%;   Score 45.4;   DB 7;   Length 168516;
Best Local Similarity   45.6%; Pred. No. 0.029;
Matches   160;   Conservative   0;   Mismatches 191;   Indels   0;   Gaps   0;

```
Qy    125  GGAGGACATGGGACTACAGTCAATGGAGATGTGTTTCAGGAGGCTAACGGCCTGCCGA  184
           ||||
Db  68477  GGAGTAAGAGGTGGAAGGGAGGCGAAGGGAGGTAGGAGGGGAGGAGGGGGGA  68536

Qy    185  TGGCTACGCTGCGATTGCCCAGGGCGACAGGTTGACTCAGGAGCCTCAGGAGCATCCGAA  244
           ||||
Db  68537  GGAGTAAGAGGTGGAAGGGAGGCGAAGGGAGGTAGGAGGAGGAGGGGGAG  68596

Qy    245  GTTGAGAGAGAGCAGAAGAAAAGGCTGCAGGAGTTGGATGCTGCCTCGAAGGTGACCGA  304
           |||| |||| |
Db  68597  GGAGGAGGAGGAGGGGGAGGAAGGGAGGAAGGGAGGTCGAAGGGAGGAGGGGGGA  68656

Qy    305  ACAGGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAACCAGCGCCAAAGTGA  364
           ||||
Db  68657  GAAGAAGGAGGAGGAGTAGGAGGAGAATAGGAGGGGGAGGAGGAGGAGG  68716

Qy    365  ACAGGTTGAGAAGAACAAGATCAACAACAGGGGGAGGAGGAAGAAGTAGGAGGAGAAGGG  424
           |||  ||
Db  68717  GGAGGAAAAGGGGGGAGGAGGAAGAAGGGAGGAGGAGGGGGAGGAGGAAGAAGGG  68776
```

```
Qy    425 CAAGGAGGAGACCCCGGCACAGAGTTGGGAGAAGGTTGGCCCAGCTGTGTGA 475
          ||||| ||        ||||||||||||||||||||||||||
Db  68777 GGAGGAGTAGGAGGGGAGGAGGAGGAGGAGGGGAGGGGGAGGGGGAGGAAGGA 68827

RESULT 13
US-11-136-527-7472/c
; Sequence 7472, Application US/11136527
; Publication No. US2005287570A1
; GENERAL INFORMATION:
;   APPLICANT: Mounts, William M
;   TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
;   FILE REFERENCE: 031896-041000 (AM101086)
;   CURRENT APPLICATION NUMBER: US/11/136,527
;   CURRENT FILING DATE:  2005-05-25
;   PRIOR APPLICATION NUMBER: US 60/574,294
;   PRIOR FILING DATE: 2005-05-26
;   NUMBER OF SEQ ID NOS: 362830
;   SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7472
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
US-11-136-527-7472

  Query Match             8.4%;  Score 45.2;  DB 7;  Length 1400;
  Best Local Similarity   50.0%;  Pred. No. 0.01;
  Matches  113;  Conservative   0;  Mismatches  113;  Indels   0;  Gaps   0;

Qy    233 GAGCATCCGCAAGTGGACTGAGAGCGAGGAGAAAAAGCTGCGAGAGTTGATGCTGCCTC 292
          ||| ||        |       |||||| | |||||||||||| ||||||
Db    533 GAGGAGGAAGAGCAGAAGAGGAAGCACGAGAAGAAGCAGGAAGGAAGAAGAAGAA 474

Qy    293 GAAGGTTGACCGAACAGAGTGGCGGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCA 352
          |||||       |       |||||||   ||||||||||
Db    473 GAAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA 414

Qy    353 GCGCCAAAGTGACCAGAGCATCAACACACAGGGCATCGGAGGAGGCTTT 412
          ||        ||||||                       ||||||
Db    413 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 354

Qy    413 TGTGAAAGAATCCAAGGAGGAGACCCCGGCACAGAGTTGGGAGAAG 458
          ||        |       |||||||||||
Db    353 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 308

RESULT 14
US-11-136-527-3376/c
; Sequence 3376, Application US/11136527
; Publication No. US2005287570A1
; GENERAL INFORMATION:
;   APPLICANT: Mounts, William M
;   TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
;   FILE REFERENCE: 031896-041000 (AM101086)
;   CURRENT APPLICATION NUMBER: US/11/136,527
;   CURRENT FILING DATE:  2005-05-25
;   PRIOR APPLICATION NUMBER: US 60/574,294
;   PRIOR FILING DATE: 2005-05-26
;   NUMBER OF SEQ ID NOS: 362830
;   SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3376
;   LENGTH: 2380
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
US-11-136-527-3376

  Query Match             8.4%;  Score 45.2;  DB 7;  Length 2380;
  Best Local Similarity   50.0%;  Pred. No. 0.011;
  Matches  113;  Conservative   0;  Mismatches  113;  Indels   0;  Gaps   0;
```

```
Qy    233 GAGCATCCGCAAGTGGACTGAGAGCGAGGAGAAAAAGCTGCGAGAGTTGATGCTGCCTC 292
          ||| |||
Db   1513 GGAGGAGGAAGAAGAAGAAGAAGAAGAAGAAGCAGAAAAAGAAGAAGAAGAAGAAGAA 1454

Qy    293 GAAGGTGACCGAACAGAGTGCGCGGAGAAGGCCAAAAAAGACCTGGAGGAGTGGAACCA 352
          ||||||                      ||||||||||
Db   1453 GAAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGCAGCAGCA 1394

Qy    353 GCGCCAAAGTGACCAGTTGAGAAGGCCATCGGAGGAGGCTTT 412
          ||        ||||||                  |||||
Db   1393 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGAAGAAGAAGAAGAGA 1334

Qy    413 TGTGAAAGAATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAG 458
          ||        |       |||||||| |||
Db   1333 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 1288

RESULT 15
US-10-821-234-49
; Sequence 49, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;   APPLICANT: Labat, Ivan
;   APPLICANT: Stache-Crain, Birgit
;   APPLICANT: Andarmani, Susan
;   APPLICANT: Tang, Y. Tom
;   TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;   FILE REFERENCE: 821A
;   CURRENT APPLICATION NUMBER: US/10/821,234
;   CURRENT FILING DATE:  2004-04-07
;   PRIOR APPLICATION NUMBER: US 60/462,047
;   PRIOR FILING DATE: 2003-04-07
;   NUMBER OF SEQ ID NOS: 1704
;   SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 49
;   LENGTH: 5468
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-821-234-49

  Query Match             8.2%;  Score 44;  DB 6;  Length 5468;
  Best Local Similarity   45.6%;  Pred. No. 0.029;
  Matches  155;  Conservative   0;  Mismatches 185;  Indels   0;  Gaps   0;

Qy    120 GGTTCGGAGGACATGGGACTACAGTCAATGGAGATGTTGTTCAGGAGGTGTTTCAGGAGGCTAACGGGCCT 179
          ||     ||   |      |       |||||||||  |||||
Db   1779 GGCAAGAAGGCCGAGGGGGACCCCGAACCAGGCAAGAAGGCCAAGAAGGCCCAGAACCAG 1838

Qy    180 GCCGATGGCTACGCTGCGATTGCCCAGGGGCACAGTTGACTCAGGAGCATC 239
          |||       |      |||                   ||||||
Db   1839 GGCAAGAAGGCCGAGGGGGACCCCGAACCAGGCAAGAAGGCCCAGAACCAG 1898

Qy    240 CGGAAGTGGAGAGAGGCGAGCAGAGAAAAGGCTGCAGGGCCCAGAACCAGGGGGCCAACCAG 299
          ||||||||||||||
Db   1899 GGCAAGAAGGCCGAGGGGGACCCCGAACCAGGCAAGAAGGCCCAGAACCAG 1958

Qy    300 ACCGAACAGAGTGCGCGGAGAAGGCCAAAAAAGACCTGGAGAGTGGAACCAGCGCCAA 359
          ||||||||    |                          ||||||
Db   1959 GGCAAGAAGGCCCAGAACCAGGCAAGAAGGCCCAGAACCAG 2018

Qy    360 AGTGAACAGGTTGAGAAGAACAAGATCAACAACACAGGGCATCGGAGGAGG 419
          ||        |       |||||||||
Db   2019 GGCAAGAAGGCCCAGAACCAGGCAAGAAGGCCCAGAACCAG 2078

Qy    420 GAATCCAAGGAGGAGACCCCAGGCACAGAGTGGGAGAAGG 459
          ||        |       |||||||| |||
Db   2079 GGCAAGAAGGCCGAGGGGGCCCAGAACCAGGGCAAGAAGG 2118
```

Search completed: January 5, 2006, 21:54:26
Job time : 315 secs